GENE EXPRESSION IN HIGHER PLANT MITOCHONDRIA

by

VALERIE P. JONES

DOCTOR OF PHILOSOPHY

UNIVERSITY OF EDINBURGH

1984

I declare that this thesis was composed by myself and that the work presented, unless otherwise stated, is my own.

Valerie Jones

October 1984

CONTENTS Page No.

Page No.

CHAPTER 4 CHARACTERISATION OF THE TRANSCRIPTS OF THE MAIZE MITOCHONDRIAL GENOME

Page No.

CHAPTER 6 DISCUSSION

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor, Dr. C.J. Leaver, for the myriad of things he has taught me about being a scientist throughout my three years at Edinburgh.

I am also very grateful to Professor L.A. Grivell and his staff at the University of Amsterdam who made my visit there in March 1984 very enjoyable and stimulating.

My thanks also go to my colleagues at the University of Edinburgh, in particular Raya Al-Shawi, John Clark, Adam Dawson, Sue Dunbar and Pete Isaac for their practical advice and helpful discussions and to my Mother for typing this thesis.

I cannot do justice here to the years of encouragement and support I have had from my parents, Kathryn and Adam.

This research was financed by an S.E.R.C. postgraduate studentship, supplemented with a travel award from the S.E.R.C.

ABSTRACT

The aim of the work presented in this thesis was to increase our understanding of the expression of the higher plant mitochondrial genome. Four particular aspects were studied.

1) In order to investigate the coding capacity of the higher plant mitochondrial genome, attempts were made to develop an *in vitro* system for the transcription and translation of fragments of mt DNA. Preliminary experiments however showed that an *E. coli* system fails to synthesise protein when primed with mt RNA.

2) Transcripts of the *Zea mays* mitochondrial genes encoding cytochrome oxidase subunits I and II (*COI* and *COII* respectively), apocytochrome *b* (*COB*) and F_1 ATPase α subunit were characterised by Northern blotting using gene-specific DNA probes. Identification of these transcripts showed that the genes (previously identified by sequence analysis only) are expressed in the mitochondrion. They also represent the first plant mt protein coding gene transcripts to be unequivocally identified and characterised.

Several interesting and unusual features have emerged from the study of mt gene transcripts. These are:

a) The transcripts of all four genes are longer than required to encode the corresponding polypeptides

b) *COI* specific probes hybridise to two major transcripts of 2.4 and 2.3 Kb. Nuclease S1 mapping has revealed that the two transcripts have 5' leaders of *ca* 152 and 57 bases respectively. The sequence identified around the start of the 2.4 Kb transcript shows 8/9 homology to the nonanucleotide sequence thought to be involved in transcription initiation

of S. cerevisiae mt DNA. It also shows 11/13 homology to a sequence around the start point of transcription of a Beta vulgaris 1.44 Kb mt DNA 'plasmid'. Homology has also been found to a sequence around the start of the Triticum aestivum COII transcript. The sequence around the start of the 2.3 Kb transcript shows 6/12 homology to the dodecamer sequence thought to be involved in processing of S. cerevisiae mt RNAs.

c) COII specific probes hybridise to a complex array of transcripts. A probe specific for the 794 bp intron hybridises to an RNA species of ca 800 bases suggesting the intron is excised from pre mRNA(s) and persists in the mitochondrion. In collaboration with A. Arnberg, circular RNA molecules of ca 800 bases have been detected in EM preparations of mt RNA. The identity of these circles has not been established.

d) COB specific DNA probes also hybridise to a complex array of transcripts. Preliminary mapping experiments suggest that the high molecular weight transcripts identified may be processed sequentially to generate the mature mRNA. Alternative explanations for the multiple transcripts of the intron-free COB gene are also discussed.

e) Transcripts of Zea mays COI, COII, and F_1 ATPase α subunit have been studied in C S and T (male sterile) type mitochondria. Most notably COII transcripts differ in male sterile cytoplasms and from N.

f) Sequences homologous to COII have been detected elsewhere in the maize mt genome.

3) Transcripts homologous to the episomal DNAs S1 and S2 have been detected. The molecular weights of the transcripts identified vary according to the presence or absence of free episomal DNAs and of copies integrated into the 'mainband' mt DNA.

4) Preliminary studies on the control of gene expression during mt biogenesis in V. faba seedlings were undertaken.

ABBREVIATIONS

A_N	Absorbance at Nnm wavelength
AMPS	Ammonium persulphate
ATA	Aurintricarboxylic acid
ATP	Adenosine 5' Triphosphate
bp	Base pairs
BSA	Bovine serum albumin Fraction V
butyl PBD	2-(4'-Butylphenyl)-5-(4"-biphenylyl)-1,3,4-oxadiazole
CIP	Calf Intestinal phosphatase
COB	Mitochondrial gene encoding apocytochrome b
COI, COII, COIII	Mitochondrial genes encoding cytochrome oxidase subunits I, II and III
cms	cytoplasmic male sterility
cpm	counts per minute
d(A,G,C and T)TP	2' deoxy (Adenosine, Guanosine, Cytidine and Thymidine) 5' triphosphates
dd(A,G,C and T)TP	2', 3' deoxy(Adenosine, Guanosine, Cytidine and Thymidine) 5' triphosphates
DCCD	N, N' Dicyclohexylcarbodiimide
DBM	Diazobenzyloxymethyl
DNA	Deoxyribonucleic acid
(r and t) DNA	DNA encoding rRNA and tRNA
dpm	disintegrations per minute
DPT	diazophenylthioether
ds	double stranded
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethyleneglycol bis (β amino ethyl ether) N, N' tetraacetic acid
EM	Electron microscope

EM	Electron microscope
FDE	Ficoll-dye-EDTA (defined section 2.2.4.1)
FSB	Formamide sample buffer (defined section 2.2.4.1)
$\underline{g}_{ave}$	Average relative centrifugal force
$\underline{g}_{max}$	Maximum relative centrifugal force
HMW	High molecular weight
IPTG	Isopropyl thio β D-galactoside
Kb	Kilobase pairs
LMW	Low molecular weight
$\log_{10}$	Logarithm (base 10)
LTB	Low Tris buffer (defined section 2.2.11.2)
M Dal	Mega Dalton (10^6 Daltons)
mA	Milliamperes
Me_2SO	dimethyl sulphoxide
MES	2-(N-morpholino) ethanesulphonic acid
MOPS	Morpholinopropanesulphonic acid
M_r	Relative molecular weight
mt	mitochondrial
NAD^+	Nicotinamide adenine dinucleotide
NDS	Naphthalene 1.5 disulphonic acid
NaPi	A mixture of NaH_2PO_4 and Na_2HPO_4 in varying proportions to generate the required pH
OD	Optical density
ORF	Open reading frame
<u>ori</u>	Origin of replication
PAS	4 amino salicylic acid
PEG	Polyethyleneglycol
PEP	Phosphoenolpyruvate
psi	Pounds per square inch
PIPES	Piperazine-N N' bis [2 ethanesulphonic acid] 1,4 Piperazinediethane sulphonic acid

PVP	Polyvinylpyrollidone
Rf	Restorer genes in maize nuclei
RF	Replicative form
RNA; (m,t and r)RNA	Ribonucleic acid (messenger, transfer and ribosomal) RNA
rpm	revolutions per minute
RuBP	Ribulose 1,5 diphosphate
RuBPCase	Ribulose 1,5 diphosphate carboxylase
S	Svedberg unit
SDS	Sodium dodecyl sulphate
ss	Single stranded
SSC	Standard saline citrate (defined section 2.1.5)
TAE	Tris acetate EDTA buffer (defined section 2.2.8.1)
TBE	Tris borate EDTA buffer (defined section 2.1.5)
TEMED	N,N,N',N' tetramethylethylene diamine
T_m	Melting temperature
TNS	Triisopropylnaphthalene sulphonic acid
Tricine	N-tris (hydroxymethyl) methyl glycine
Tris	Tris (hydroxymethyl) aminomethane
Triton X-100	Octylphenoxypolyethoxy ethanol
URF	Unassigned reading frame
UV	Ultra violet
v/v	volume per volume
w/v	weight per volume
X gal	5-bromo 4-chloro 3-indolyl β D-galactoside

CHAPTER 1 GENERAL INTRODUCTION

1.1 INTRODUCTORY REMARKS

Biogenesis of a functional mitochondrion is dependent on the co-ordinate expression of nuclear and mitochondrial genes. The nature of the genetic information contained in the mitochondrial DNA (mt DNA) of many organisms and its interaction with the nuclear genome have been studied extensively in recent years, and an understanding of some aspects of their interplay is now being reached. However, studies on the information content and expression of plant mitochondrial genes has lagged far behind those on mammals and fungi, and the overall aim of the work described in this thesis is partly to redress that balance.

The purpose of this chapter is to present an up-to-date account of the salient features of mitochondrial gene organisation and expression, with emphasis on plants and in particular maize (Zea mays L).

1.2 GENERAL CHARACTERISTICS OF MITOCHONDRIA

Mitochondria are enveloped by two distinct membranes, the highly invaginated inner membrane being the site of electron transport, ATP synthesis and ion transport. The matrix, completely surrounded by the inner membrane, contains the soluble enzymes, the DNA, and the protein synthetic machinery. Mitochondria can account for between 5 and 20% of the protein in a eukaryotic cell and are thus major cellular constituents (Schatz 1983). There is great variation in the number of individual mitochondria per cell. The extremes are represented by Saccharomyces cerevisiae grown under certain conditions which contain 1 - 3 mitochondria per cell (Stevens 1981) and by

amphibian oocytes which contain in the order of 10^7 mitochondria (Schatz 1983). A typical mammalian hepatocyte contains between 1,000 and 2,000 mitochondria (Bogenhagen and Clayton 1974). The number and morphology of mitochondria in a cell can vary according to the physiological conditions. For example yeast cells grown under high glucose conditions contain a few highly convoluted mitochondria, whereas cells grown on low glucose levels contain up to 50 ellipsoid mitochondria (Stevens 1981).

Mitochondria can interact with each other in the cell. Examples include the fusion of mitochondria to form a single spiral organelle during spermatogenesis (Munn 1974), and fusion of mitochondria after mating of yeast cells (Strausberg and Butow 1977).

Mitochondria have been shown to contain their own genetic system, distinct from that of the nucleus (Luck 1963), but are in no way autonomous. Only a small fraction of the total number of organelle proteins is encoded within the mitochondria themselves; all other proteins are specified by nuclear genes, synthesised in the cytoplasm, and imported. It is not known why mitochondria and chloroplasts contain their own genetic systems while other organelles such as peroxisomes do not.

1.3 MITOCHONDRIAL DNA: SIZE AND PHYSICAL CONFORMATION

1.3.1 General characteristics

Mitochondrial DNAs are diverse in size and physical form. Circular forms predominate, but linear mt DNAs are found in the protozoans Tetrahymena (Suyama and Miura 1968) and Paramecium (Goddard and Cummings 1975) and some fungi eg. Hansenula mrakii (Wesolowski and

Fukuhara 1981). In the protozoans Trypanosoma and Crithida, mt DNA is enmeshed in networks containing maxi- and mini-circular DNAs, the latter of unknown function (Borst and Hoeijmakers 1979). In this chapter only the organisation of animal, fungal and plant mt DNAs is discussed.

1.3.2 Animal mitochondrial DNAs

The mt DNAs of animals are among the smallest functional mt genomes identified. Electron microscopic analysis of mt DNA from a variety of animals reveals the presence of double-stranded, circular, super-coiled DNA molecules with a contour length of 4.85 - 5.85 μm (reviewed in Borst 1972). These size estimates agree closely with those obtained by complete sequence analysis of the human (16.56 Kb, Anderson et al 1981), bovine (16.33 Kb, Anderson et al 1982) and mouse (16.29 Kb, Bibb et al 1981). The mt DNAs from Xenopus laevis (Dawid 1972) and Drosophila melanogaster (Fauron and Wolstenholme 1980) have sizes similar to those in mammals, although they have not been completely sequenced.

1.3.3 Fungal mitochondrial DNAs

Fungal mt DNAs from a variety of species show great diversity in their size, ranging from ca 19 Kb in Torulopsis glabrata (Clark-Walker et al 1981) and Schizosaccharomyces pombe (Anziano et al 1983) to 108 Kb in Brettanomyces custersii (Clark-Walker et al 1981, McArthur and Clark-Walker 1983). In general, estimates of genome size based on electron microscopy and summation of restriction enzyme fragments are in close agreement for fungal mt DNAs. Using these approaches, the mt genome of S. cerevisiae has been shown to exist as a 74 - 82 Kb circular molecule, (Hollenberg et al 1970), 70% of which

has been sequenced (L. Grivell pers. commun.) The S. cerevisiae mt genome contains long intergenic stretches of A + T-rich DNA within which are short G + C clusters (Bernardi and Bernardi 1980, Cosson and Tzagoloff 1979). The function, if any, of these A + T-rich 'spacer' regions is unknown; however, since many yeast mt transcripts are initiated within these regions the term 'spacer' DNA may be inappropriate.

Complete sequence analysis of two other fungal mt genomes namely Schizosaccharomyces pombe (19 Kb, Lang et al 1983) and Aspergillus nidulans (33 Kb, Scazzocchio et al 1983) revealed no A + T-rich regions; neither are they found in the mt DNA of Podospora anserina (95 Kb, Wright et al 1982) or of Neurospora crassa (61 Kb, Burke and RajBhandary 1982).

1.3.4 Plant mitochondrial DNAs

Until very recently there have been many problems in estimating the complexity and physical conformation of the higher plant mt genome. Only two features were generally agreed: (i) its remarkably uniform buoyant density from a wide range of plants (1.705 - 1.707 g/cm^3, equivalent to a G + C content of 47%, Wells and Ingle 1970, Shah and Levings 1974) except for Oenothera mt DNA which has a density as high as 1.710 g/cm^3 equivalent to G + C = 51% (Brennicke, 1980) and (ii) that it was probably larger than the yeast mt genome. Initial electron microscopic studies on the mt genome of pea revealed that the majority of mt DNA could be isolated as circles of contour length 30 - 35 μm, equivalent to 70 Mdal or 100 Kb (Kolodner and Tewari 1972). These results were in close agreement with the estimates of genome size (74 Mdal) based on reassociation kinetics (Kolodner and Tewari 1972).

More recent estimates, however, made by summation of restriction fragment sizes and by renaturation kinetics suggest that the mt genome of pea may be more complex (200 - 250 Mdal, Ward et al 1981). Many other studies based on restriction enzyme analysis and restriction fragment mapping have shown that generally the mt genomes of higher plants are larger and more complex than indicated by Kolodner and Tewari. These values are summarised on Table 1.1

Table 1.1 Size estimates of some plant mt DNAs

Species	Size Kbp	Basis of Estimate	Reference
Pisum sativum	100	b + c	Kolodner and Tewari 1972
	360	a + c	Ward et al 1981
Brassica campestris	218	d	Palmer and Shields 1984
Cucurbitaceae	320-2,400	a + c	Ward et al 1981
Zea mays	480	a + c	Ward et al 1981
	570	d	Lonsdale et al 1983
Oenothera berteriana	185	a	Brennicke 1980
Triticum aestivum	430-440	d	F. Quetier pers. commun.
	210	a	Quetier and Vedel 1977
Solanum tuberosum	135	a	Quetier and Vedel 1977
	90	b	Quetier and Vedel 1977
Glycine max	90	b	Synenki et al 1978
Parthenocissus tricuspidata	240	a	Quetier and Vedel 1977

a summation of restriction fragment sizes

b EM analysis

c reassociation kinetics

d restriction mapping

In general, the genome sizes calculated by restriction enzyme analysis are consistentlyhigher than EM measurements. This is particularly marked for maize, where summation of restriction fragments indicated a complexity of between 180 and 320 Mdal (Ward et al 1981) whereas EM analysis of mt DNA revealed three main size classes of circular molecules of 33, 45 and 66 Mdal (Levings et al 1979) [although the majority of mt DNA was linear]. In order to reconcile these anomalies two alternative models have been proposed:

1. The plant mt genome consists of groups of molecules carrying identical information, but which are partially methylated, thus increasing the number of restriction fragments obtained with methylation sensitive enzymes. However, the use of isoschizomeric restriction enzymes differing in their ability to cleave methylated DNA has shown this to be unlikely (Bonen et al 1980, Ward et al 1981).

2. The plant mt genome is distributed over physically indistinguishable molecules or "chromosomes" (Quetier and Vedel 1977, 1980, Spruill et al 1980). However, these authors offer no explanation as to how complete mt genomes are transmitted to daughter mitochondria. They also base their models on a highly selected population of circular mt DNA molecules; the majority of plant mt DNA as isolated is linear (Quetier and Vedel 1980).

These inconsistencies in the organisation of the plant mt genome have largely been circumvented by the construction of restriction maps using cosmid or plasmid clone banks of mt DNA. Palmer and Shields (1984) have constructed a complete restriction map of the Brassica campestris mt genome, which has been shown to exist as a circular molecule of 218 Kb. Within this "master chromosome", which contains

the entire sequence complexity of the mt genome, are two direct repeats of 2 Kb separated by 135 and 83 Kb. Evidence for intramolecular recombination across this repeat comes from analysis of restriction sites around the 2 Kb repeat; four PstI fragments containing the 2 Kb repeat can be isolated, and these are shown to be flanked by four paired combinations of four unique sequences. Moreover the 2 Kb repeat sequences characteristic of the small circles are present at twice the molar concentration of those of the "master chromosome". Thus the Brassica compestris mt genome is likely to exist largely as a tripartite structure with equimolar amounts of the small circles present at twice the molar concentration of the "master chromosome". Stern and Palmer (1984a) have shown that the 2 Kb repeat involved in recombination of the Brassica mt genome contains or is closely linked to the COII gene.

There is now an increasing body of evidence to suggest that other plant mt DNAs exist as discrete circular chromosomes maintained by recombination across repeated element(s). Restriction mapping of cosmid clones of maize mt DNA suggests it exists as a master circular chromosome of 570 Kb which contains at least six pairs of inverted and direct repeats, and recombination has been shown to occur across four of these (Lonsdale et al 1983, T. Hodge pers. commun.). Intramolecular recombination across two direct repeats of 12 Kb separated by 67 Kb leads to the formation of two circles of 67 and 503 Kb. Recombination across two 3 Kb direct repeats within the larger molecule generates circles of 250 and 253 Kb (Lonsdale et al 1983). The latter event is thought to occur at high frequency since the four restriction fragments containing the 3 Kb repeat (and hence the three circles) are present in equimolar amounts. Recombination across other

repeated elements in the Zea mays mt genome are thought to be relatively rare; however by modelling recombinational events, it is possible to generate size classes of molecules which agree closely with those visualised by EM (D. Lonsdale pers. commun.).

The wheat mt genome contains 10 repeated elements on a master chromosome of 430 Kb. However unlike Zea mays or Brassica campestris, hybridisation studies have shown that two of these repeated elements contain the rRNA genes. One 4 Kb repeat contains the 18S + 5S rDNAs and another repeat (5 Kb) contains the 26S rDNA (Falconet et al 1984, B. Lejeune pers. commun.). The 26S rDNA is also closely associated with repeated sequences in spinach and pokeweed mt DNA (Stern and Palmer 1984a) implying that association between recombination repeats and plant mt genes may be a common occurrence. Figure 1.1 shows the way in which recombination across a pair of direct repeats in a single circular molecule gives rise to two smaller circles. Thus it appears that the model of a single master chromosome which is able to undergo intramolecular recombination(s) may apply to all higher plant mitochondria. If so, it goes some way to explaining why the bulk of mt DNA observed by electron microscopy is linear; circles of ca. 500 Kb would be difficult to isolate intact.

Such intramolecular recombination is found to occur in a number of other extranuclear genetic systems, including the chloroplast genomes of Phaseolus vulgaris and Osmunda (Palmer 1983) and the 2μ plasmid of yeast (Broach 1981) where rapid recombination across inverted repeats generates equimolar populations of "flipped" molecules.

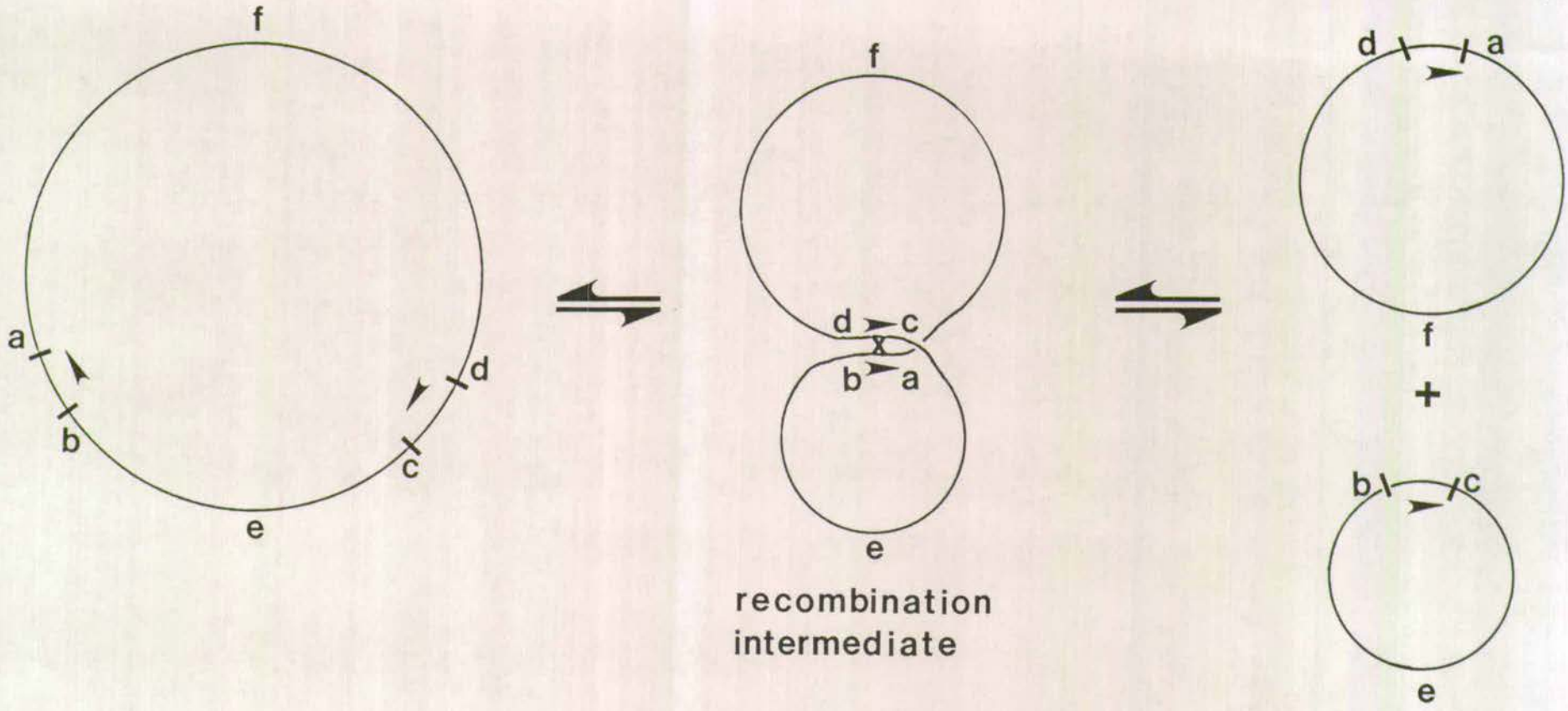


Fig 1.1 Recombinational model for the higher plant mt genome

Recombination is proposed to take place across a pair of direct repeats ◄ flanked by sequences a, b, c and d. Recombination across the direct repeats generates novel orientations of the flanking sequences with respect to the repeat. Note that intramolecular recombination across direct repeats generates two equimolar smaller molecules, cf recombination across an inverted repeat. In Brassica campestris, e = 83 Kb, f = 135 Kb and ◄ = 2 Kb. In Zea mays e = 250 Kb, f = 253 Kb and ◄ = 3 Kb.

1.4 MITOCHONDRIAL GENES AND GENE ORGANISATION

1.4.1 Mitochondrial genes: general characteristics

All mt DNAs characterised to date encode components of the mitochondrial translation system and a limited number of proteins forming subunits of the respiratory enzyme complexes of the inner mitochondrial membrane. Genes identified in the yeast, mammalian and plant mt genomes are listed in Table 1.2. Gene complements are similar, but there is great variation in their organisation and expression.

1.4.2 Animal mitochondrial genes

Complete sequence analysis of the human (Anderson et al 1981), bovine (Anderson et al 1982) and mouse (Bibb et al 1981) mitochondrial genomes has revealed a striking conservation of gene content, organisation and even sequence. Each encodes 12S and 16S rRNAs, 22 tRNAs and 13 open reading frames (ORFs). All but 1 ORF and 8 tRNAs are transcribed from the same strand (the Heavy (H) strand). Coding sequences are immediately contiguous or are separated by only a few nucleotides, and in most cases are flanked on either side by a tRNA gene. It has been suggested that these tDNAs may act as 'punctuation signals' for enzymes involved in processing of a primary polycistronic transcript from the 'H' strand (Ojala et al 1981). Overlapping coding sequences have been found in mammalian mt DNAs; for example the genes encoding ATPase 8 and ATPase 6 overlap by 46 nucleotides in different reading frames. None of the genes contains introns. Of the 13 ORFs, four have been identified by comparison with the amino acid sequence of mammalian mitochondrial proteins (cytochrome c oxidase subunits I, II and III, and apocytochrome b; COI, COII, COIII and COB respectively). Comparison with

Table 1.2 Mitochondrial genes in plants, yeast and mammals

Mitochondrial component	mitochondrial gene product in plants	*S. cerevisiae*	mammals
Ribosomes			
large ribosomal RNA	26S	21S	16S
small ribosomal RNA	18S	15S	12S
5S ribosomal RNA	+	-	-
ribosomal associated protein	+	+	-
ATPase complex			
subunit 6	?	+	+
subunit 8	+	+	+
subunit 9	+	+	-
α subunit F_1	+	-	-
cytochrome *c* oxidase			
subunit I	+	+	+
subunit II	+	+	+
subunit III	?	+	+
Ubiquinol - cytochrome *c* reductase			
apocytochrome *b*	+	+	+
RNA maturases	?	+	-
tRNA synthesis locus	?	+	-
URFs	at least 1	6*	7**
tRNAs	?	25	22

* mostly intronic
** no homology with yeast URFs

Modified and adapted from Grivell, 1983

yeast gene sequences has also allowed the genes encoding ATPase 6 (Macino and Tzagolaff 1980) and ATPase 8 (Macreadie et al 1983) to be identified. The identity of the remaining seven ORFs is not known (designated unassigned reading frames or URFs), although all except URF6 are transcribed (Montoya et al 1981, Ojala et al 1981) and at least one (URF3) is transcribed and translated (Oliver et al 1983). The identification of some of these URFs in the mitochondrial genomes of other organisms for example maize (URF-1, Brown et al 1983), Aspergillus nidulans (URFs 1 and 4, Brown et al 1983) and Neurospora crassa (URFs 1 and 5, Nelson et al 1983), suggests they may play an important general role in mitochondrial biogenesis.

Comparison of the predicted amino acid sequence of mammalian mitochondrial genes with the corresponding polypeptide sequences (Barrell et al 1979) reveals that mammalian mitochondria use a novel genetic code, most notably UGA encodes tryptophan (normally termination) and AUA specifies methionine (normally isoleucine).

1.4.3 Fungal mitochondrial genes

The Saccharomyces cerevisiae mitochondrial genome (74-82 Kb) is five-fold larger than its mammalian counterpart, and yet retains a very similar, though not identical, gene complement (Table 1.2). The most notable differences are the presence of the genes encoding the ribosome associated protein, var-1 (Groot et al 1979) and F_o ATPase subunit 9 (Hensgens et al 1979) and a few short ORFs of unknown function (Coruzzi et al 1981). No URFs homologous to the mammalian URFs have been discovered.

The difference in size relative to the human mt genome can almost entirely be accounted for by differences in gene organisation. Firstly,

the yeast mt genome contains long stretches of A + T-rich sequences (Section 1.3.3). The majority of this 'spacer' DNA does not contain polypeptide coding sequences with the exception of var-1 (Zassenhaus and Butow 1983). Secondly, three of the genes in S. cerevisiae (COI, COB and 21S rRNA)are split by one or more introns (Borst 1979, Hensgens et al 1983, Labouesse and Slonimski 1983). Many of these introns are "optional" in that they are present in some strains but not others. For example the 'long' form of COB contains five introns (strains 777-3A and KL 14-4A, Sanders et al 1977), whereas the 'short' form lacks three of these (strain D273-10B, Labouesse and Slonimski 1983). Some yeast mt introns contain open reading frames which encode proteins (dubbed maturases) which may mediate splicing of the introns that encode them (Lazowska et al 1980).

The organisation of the yeast mitochondrial genome differs from that of mammalian mitochondria. Most notably the tRNA genes are mostly clustered and not interspersed with the protein coding genes (Borst and Grivell 1981). The genetic code used by S. cerevisiae mitochondria differs from both the 'universal' code and that used by mammalian mitochondria. Like mammals UGA specifies tryptophan (Fox 1979) and AUA specifies methionine (Hudspeth et al 1982). However, CU (N) where N is any nucleotide specify threonine rather than leucine (Li and Tzagoloff 1979).

Within the fungi themselves there are considerable differences in the organisation of mitochondrial genetic information despite similarities in gene complements. The only major differences in coding capacity are the presence of species-specific URFs in A. nidulans and N. crassa (section 1.4.2), and the presence of an active gene encoding F_o ATPase subunit 9; some fungal mitochondrial genomes contain

apparently 'silent' copies of this gene (section 1.10). More detailed reviews of the mitochondrial DNA of N. crassa, A. nidulans and Podospora anserina can be found in Burger and Werner (1983), Scazzocchio et al (1983) and Wright and Cummings (1983) respectively. The organisation of one of the smaller fungal mitochondrial genomes (that of Schizosaccharomyces pombe, Lang et al (1983)) is similar to its mammalian counterparts, with tRNA genes flanking many coding sequences.

1.4.4 Plant mitochondrial genes

Unlike the mammalian and fungal mitochondrial genomes, the information content and gene organisation of plant mitochondrial DNAs have not been investigated until recently. This is due partly to their large size, so preventing direct sequencing as for the mammalian mitochondrial genomes, and also to a lack of non-lethal mutations, the availability of which has greatly expedited the identification of yeast mitochondrial genes (although see Section 1.8.2). The relatively low A + T content of plant mitochondrial DNA together with its large complexity (Ward et al 1981) suggests a potential coding capacity much greater than that of the fungi. However, there is evidence (see below) that not all this extra potential coding capacity is realised.

A variety of approaches have been used to identify plant mt genes. This topic has been the subject of recent reviews (Leaver and Gray (1982), Leaver et al (1983)) and the salient points are described below.

1.4.4.1 Analysis of mitochondrial translation products

Isolated mitochondria from a variety of plants synthesise 18 - 20 polypeptides when analysed by one dimensional SDS polyacrylamide gel electrophoresis, and a further 20 - 30 can be resolved by two dimensional separation (Leaver and Pope 1976, Forde et al 1979, Hack and Leaver 1983). These polypeptides are presumed to be encoded by genes in the mt DNA because 1) their synthesis is insensitive to inhibitors of cytoplasmic protein synthesis and 2) mRNA import into mitochondria has not been demonstrated. However, these results must be interpreted with caution, since isolated human mitochondria synthesise up to 26 polypeptides even though the mt genome only has the capacity to encode 13 of them (Attardi 1981, Attardi et al 1983). These additional polypeptides may result from proteolysis and/or premature termination of translation; hence the complexity of the pattern of mitochondrially synthesised proteins does not necessarily reflect a correspondingly complex mt genome.

A variety of techniques have been used to identify tentatively polypeptides synthesised by isolated maize mitochondria.

1) Cytochrome oxidase subunits I & II have been identified by immunoprecipitation with antibodies raised against the homologous yeast polypeptides (Forde and Leaver 1979).

2) A polypeptide of M_r 8,000 is thought to be subunit 9 of F_o ATPase as it is soluble in organic solvents and because it can bind ^{14}C - DCCD (Hack and Leaver 1984).

3) A polypeptide of M_r 44,000 co-purifies with mitochondrial ribosomes and is probably analagous to the yeast Var-1 protein (Leaver et al 1982).

4) The α subunit of F_1 ATPase was identified by comigration of a mitochondrially synthesised polypeptide with purified α subunit on one- and two-dimensional SDS polyacrylamide gels (Hack and Leaver 1983). Partial proteolysis and immunoprecipitation with antibodies raised against the yeast α subunit have confirmed this identification. The α subunit has been shown to be a mitochondrial translation product in cucumber, sorghum, pea and field bean (Hack and Leaver 1983). A mitochondrial location for this gene is unexpected since in all organisms other than plants it is encoded in the nucleus.

The remaining polypeptides are, as yet, unidentified but by analogy with fungal and mammalian systems it is likely that cytochrome b, cytochrome oxidase subunit III and F_o ATPase subunits 6 and 8 will be amongst them.

1.4.4.2 Identification of genes by nucleic acid hybridisation

Another approach to the identification of genes has been by hybridisation of known,well characterised DNA and RNA molecules to the mitochondrial DNA. Hybridisation of ^{32}P labelled ribosomal RNAs (26S, 18S and 5S, Leaver and Harmey 1973, Leaver and Harmey 1976) to mt DNA has shown that the genes encoding them are found in the mitochondrial DNAs of wheat (Bonen and Gray 1980), maize (Stern et al 1982, Iams and Sinclair 1982) and Oenothera (A. Brennicke, pers. commun.). These hybridisation studies and subsequent sequence analyses have revealed that the genes for 18S and 5S rRNA are closely linked in the mt DNA [spacer region varies from 108 in maize (Chao et al 1983), 114bp in wheat (F. Quetier, pers. commun.) to ca 300 Bp in Oenothera (A. Brennicke pers. commun.)] but well separated from 26S rRNA gene (by 15 - 16 Kbp in maize). This arrangement of small

and large rRNA genes is similar to that in yeast (Borst and Grivell 1978) but differs from their organisation in mammalian mitochondria where 16S and 12S rRNAs are closely linked (Anderson et al 1981) although only plants are known to possess a 5S rRNA gene. The genes encoding the rRNAs in wheat are present as multiple copies in the mitochondrial genome (Falconet et al 1984) whereas in maize they are represented only once.

Sequence analysis of the wheat 26S (M. Gray pers. commun.), 18S (Spencer et al 1984) and 5S (Spencer et al 1981) rRNAs is now complete. A partial sequence of the maize 18 - 5S rDNA unit has been reported (Chao et al 1983)and the maize 26S rDNA has been sequenced (Dale et al 1984). Both the 26S and 18S rRNAs show pronounced structural similarities with the 23S and 16S rRNAs of E. coli.

A similar approach has been used to show that the wheat and maize mitochondrial genomes encode several tRNAs which appear to be clustered as in yeast mt DNA (D. Lonsdale pers. commun). The sequence of wheat $tRNA^{met}_{f}$ has recently been reported (Gray and Spencer 1983).

Protein-coding genes have been identified in plant mt DNAs by the hybridisation of well defined probes for mitochondrial genes from yeast and beef to restriction fragments of mt DNA under 'heterologous conditions'. The genes identified to date are:

1) COII from maize (Fox and Leaver 1981), wheat (L. Bonen pers. commun.), Oenothera (Hiesel and Brennicke 1983), rice and pea (Kao et al 1984). In maize, wheat and rice the COII coding sequence is interrupted by single introns of 794, 1,200 bp and 1,265 bp respectively.

2) COI from maize (Isaac et al 1984), Chlamydomonas (M. Gray, pers. commun.) and Sorghum (D. Hanson pers. commun.). None of these genes contain introns.

3) COB from maize (Dawson et al 1984) and Oenothera (Hiesel and Brennicke 1984) neither of which is interrupted.

4) An open reading frame showing some homology to the mammalian URF-1 has been identified and partially sequenced in maize mt DNA (Brown et al 1983).

5) F_o ATPase 8 has been identified in Oenothera (Hiesel and Brennicke 1984a) by comparison with the yeast and mammalian sequences. This gene shows an unexpected organisation in that it overlaps COII by four bases.

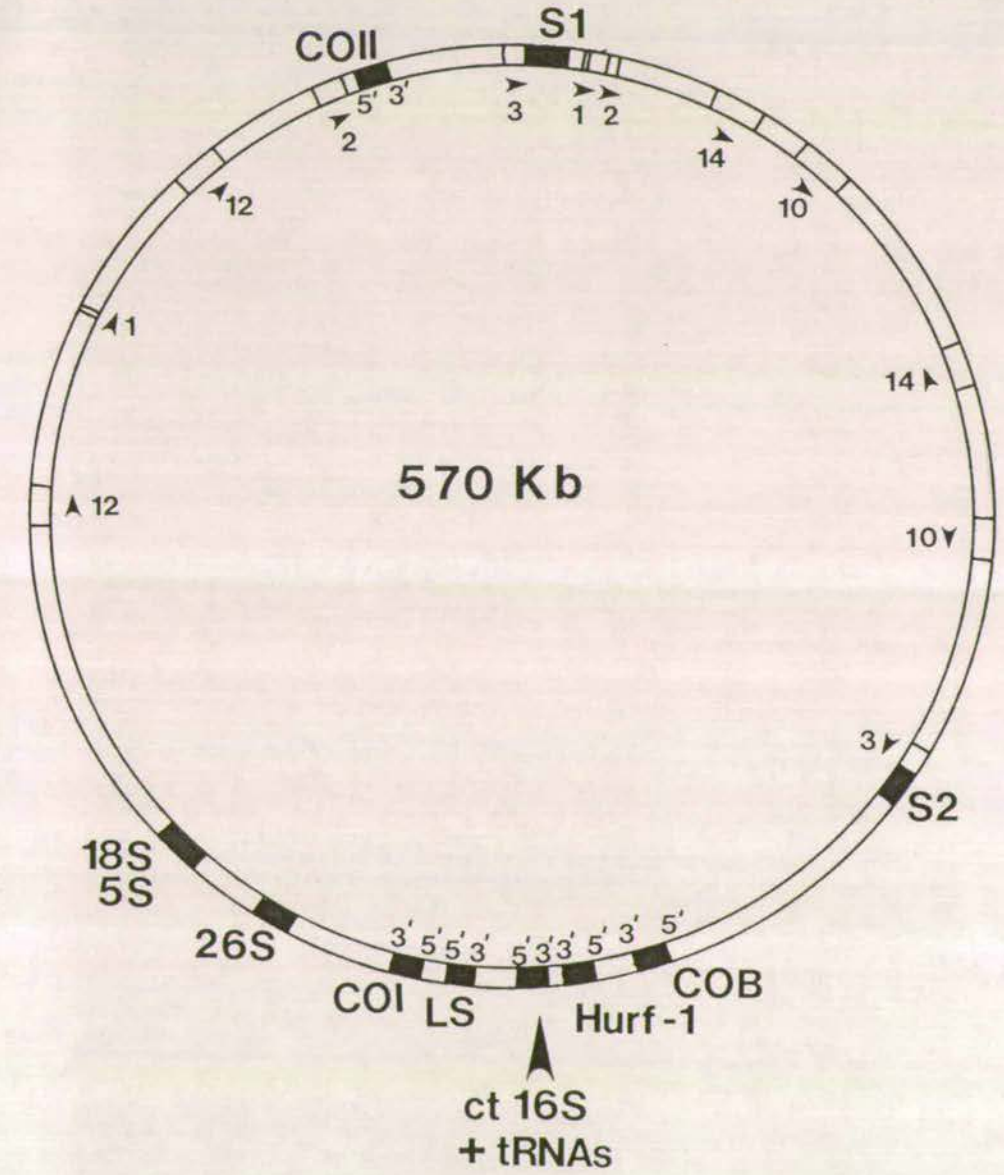
Sequence analysis of these genes and comparison of the predicted amino acid sequence with that from yeast and beef suggests that the higher plant mitochondrial genetic code differs from both the 'universal' code and the mammalian and fungal mitochondrial genetic code. In particular CGG (normally Arginine) is thought to encode tryptophan, and there is no evidence that UGA encodes tryptophan as in other mt DNAs. This codon is used as a normal termination codon in the Oenothera COB gene (Hiesel and Brennicke 1984). Most of these genes have now been located on the circular restriction map of the maize mitochondrial genome (Fig. 1.2). COI, URF-1 and COB are loosely linked in a 40Kbp region, whereas COII is over 100 Kb from COI (A. Dawson pers. commun.). All four genes appear to be transcribed from the same DNA strand.

It is clear from these studies that the maximum potential number of mitochondrial genes (predicted from the translation products of

Fig 1.2 Map of the maize mitochondrial genome

The locations of various protein and rRNA coding genes on the map of the maize mt genome has been determined by hybridisation with a cosmid bank of mt DNA clones. The episomal DNAs S1 and S2 (section 1.8.2.2) are also marked.

Drawn from D Lonsdale, pers. commun.
LS = large subunit of RuBPCase
16S = Chloroplast 16S rRNA
Hurf-1 = Human mitochondrial URF-1



3 = Repeated sequence with length (Kb)

isolated mitochondria) would only occupy a very small proportion of the higher plant mitochondrial genome. Even allowing for the presence of genes normally found in the nucleus (F_1 α ATPase) and sequences apparently derived from the chloroplasts (Section 1.10.9) much of the plant mitochondrial genome must be composed of long stretches of DNA which has a sequence independent function or even no function at all.

1.5 REPLICATION OF MITOCHONDRIAL DNA

A detailed review of the replication of mammalian mt DNA can be found in Clayton (1982). Analysis of replication intermediates separated on caesium chloride gradients has revealed that the replication of the two strands [Heavy (H) and Light (L)] begins at two points on the mt DNA. Replication of the H strand is initiated in the D loop region (a region of the mt genome formally equivalent to a DNA triplex structure) and proceeds until the L strand replication origin is displaced, at which point L strand replication begins.

Replication of yeast mt DNA differs from mammalian mt DNA replication since at least seven separate origins (ori) are involved (Bernardi 1982). Analysis of the DNA sequences retained in hyper-suppressive petites shows that the ori sequence is approximately 300 bp long and consists of three G + C clusters separated by A + T-rich stretches. This sequence can form secondary structures similar to those found at mammalian mt origins of replication (de Zamaroczy et al 1981). Since the nonanucleotide sequence thought to be involved in initiation of transcription (section 1.6.3) is found at ori sequences in these petites, Osinga et al (1982, 1984) have concluded that this sequence is involved in RNA priming of DNA replication.

Replication of plant mt DNA has not been investigated in depth and as yet no origins of replication have been found.

1.6 TRANSCRIPTION OF MITOCHONDRIAL DNA

1.6.1 Mitochondrial RNA polymerases

Mitochondrial RNA polymerases are thought to be encoded in the nuclear genome and imported into the mitochondria. Since RNA polymerase plays such a central role in the expression of mt DNA, regulation of the activity of the enzyme could be one way in which the nucleus exerts control over the mitochondrial gene expression. Mitochondrial RNA polymerase has been purified from rat liver (Mukerjee and Goldfeder 1973), Xenopus laevis (Wu and Dawid 1972), Homo sapiens (Walberg and Clayton 1983), Saccharomyces cerevisiae (Levens et al 1981) and Neurospora crassa (Kuntzel and Schafer 1971).

The purest mt RNA polymerase preparations are resolved to one band by electrophoresis on SDS polyacrylamide gels. Typical molecular weights are:

46,000 in X. laevis, 45,000 in S. cerevisiae, 64,000 in N. crassa, 66,000 in rat.

The Homo Sapiens mt RNA polymerase has only been partially purified, to give several polypeptides in the range 45,000 to 64,000.

The most highly characterised mt RNA polymerase is that of S. cerevisiae. The enzyme sediments at 6.3S in glycerol gradients which is commensurate with a molecular weight of 100,000 to 150,000 assuming a globular structure (Levens et al 1981). Thus the 45,000 M_r polypeptide cannot be the active enzyme since this would sediment between 3 and 4S. The S. cerevisiae mt RNA polymerase is therefore likely to be a

multimeric enzyme composed of identical or non-identical 45,000 M_r polypeptides. The purified mt RNA polymerase from S. cerevisiae and the partially purified human mt RNA polymerase have been used to characterise in vitro 'run off' transcripts from defined DNA restriction fragments (see for example Edwards et al 1983, Chang and Clayton 1984, Bogenhagen et al 1984). The seemingly relatively simple organisation of mt RNA polymerases from a variety of organisms contrasts with the more complex RNA polymerases of E. coli (subunits of M_r ca 160,000(β'), 150,000(β), 70,000(σ) and 36,500(γ) reviewed by Chamberlin (1982), Z. mays chloroplasts [subunits of M_r 200,000, 180,000 as well as several smaller polypeptides (Smith and Bogorad 1974)] and the nuclear RNA polymerases I II and III where each type is composed of many subunits (reviewed by Lewis and Burgess 1982). All mt RNA polymerases characterised are insensitive to α amanitin.

The biosynthesis of S. cerevisiae mt RNA polymerase has been studied by cell free translation of RNA isolated from yeast cells in different growth stages. One of the polypeptides synthesised has a M_r of 47,000 which is antigenically related to the 45,000 M_r mt RNA polymerase, and probably represents a precursor form. RNA harvested from cells in late logarithmic or stationary phase direct the synthesis of increased levels of this M_r 47,000 polypeptide relative to mid-log cells. This induction of mt RNA polymerase follows the same course as the induction of the cytochromes, and thus can be correlated with an increase in the synthesis of mitochondrially encoded proteins (Levens et al 1982).

1.6.2 Mammalian mitochondrial RNAs

The recent determination of the complete nucleotide sequence of several mammalian mt DNAs together with detailed analyses of the structural and metabolic properties of human mitochondrial trans-

cripts has allowed a fairly complete model to be proposed for the transcription of these mt DNAs (Attardi et al 1983, Ojala et al 1981, Montoya et al 1981). Both the H strand (which contains most of the genes) and the L strand (which contains 8 tRNA genes and one URF) are completely and symmetrically transcribed from promoters in the D loop region (section 1.5 , Murphy et al 1975). In spite of its limited informational content the L strand is transcribed at 2 - 3 times the rate of the H strand (Cantatore and Attardi 1980). The reason for this high rate of L strand transcription is not known. The compact organisation of the mammalian mt DNA is reflected in its pattern of transcription, the salient features of which are:

1) Mammalian mt mRNAs are not capped since the 5' sequence of the mRNA aligns exactly with the coding sequence (Montoya et al 1981).

2) The 5' of most mRNAs are immediately adjacent to, or separated by a few nucleotides from tRNA genes (Anderson et al 1981).

3) The 3' terminal residue of most mt mRNAs corresponds to a DNA residue immediately adjacent to the 5' end of a tRNA gene.

4) Mammalian mitochondrial mRNAs lack significant 5' non-coding stretches, since initiation codons are usually found within six nucleotides of the 5' terminus.

5) Mature mRNAs are polyadenylated with ca 55 nucleotides (Hirsch and Penman 1974). This is necessary to complete the termination codon in some mammalian mt genes which end with the incomplete triplets U or UA.

These features of mammalian mt transcripts are consistent with a model for H-strand transcription where a single complete transcript is processed before and after each tRNA sequence to yield mature mRNAs. Where tRNAs are not found at a putative processing site for example

between ATPase 6 and COIII, the processing enzyme may recognise an alternative secondary structure (Bibb et al 1981). Failure to detect any giant sized H strand precursors suggests processing of transcripts takes place while they are still part of the transcription complexes.

An obvious limitation of this model is that it places the expression of rRNA, tRNA and protein coding genes under the control of one promoter. While the products of the protein coding genes may be required in fairly equal stoichiometries, it does not explain the sixty-fold molar excess of the rRNAs (Attardi et al 1982). Recent experimental evidence strongly supports the existence of two separate H strand promoters. Initiation of RNA synthesis just upstream from a $tRNA^{Phe}$ gene occurs frequently and gives rise to transcripts terminating at the 3' end of the 16S rRNA. Initiation at the second promoter which is downstream of the first and located near 5' end of the 12S rRNA gene is much less frequent and results in the synthesis of a polycistronic transcript of the entire H strand. The evidence for two separate promoters is as follows:

1) Mapping of nascent transcripts on mt DNA identify two sites for the initiation of transcription: one is situated very close to the 5' terminus of the 12S rRNA gene and the other is 90 - 110 bp upstream, to the 5' of a $tRNA^{Phe}$ gene (Montoya et al 1982).

2) Kinetics of labelling experiments showed there were two types of transcription event, one giving rise to the rRNAs, the other to transcription of the entire H strand mRNAs (Attardi et al 1983, Montoya et al 1983).

3) Primary transcripts "capped" in vitro with $\alpha\ ^{32}P$ - GTP and

guanylyl transferase (section 1.6.3) also map to two distinct positions on the mt genome (Attardi et al 1983).

A consequence of this model is that the fate of a transcript is determined either by the involvement of two RNA polymerases, or by the sequences present at the 5' end of an individual transcript.

In vitro transcription of human mt DNA has identified potential promoter sequences involved in the initiation of transcription of the H and L strands (Bogenhagen and Clayton 1984, Chang and Clayton 1984). In this assay the purified mt RNA polymerase only initiates from the 'promoter' upstream of the $tRNA^{Phe}$ and 12S rRNA. Transcription is not initiated at the junction between $tRNA^{Phe}$ and 12S rRNA, postulated to be the position where mRNAs initiate. This probably reflects its in vivo role as a 'weak' promoter. The human L and H strand putative promoters share the sequence 5' CANAC$^{C}_{G}$ CCAAAGAPyA 3' (Chang and Clayton 1984) and transcription initiation occurs 6 - 8 bp 3' of this sequence.

Post-transcriptional modifications of human mt transcripts include cleavage of individual transcripts from the polycistronic mRNA, polyadenylation, and the addition of the non-coded 5' CCA 3' to the 3' end of tRNA molecules (Anderson et al 1981).

1.6.3 Fungal mitochondrial RNAs

The scope for differential transcription and post-transcriptional control of mt gene expression is far more extensive in yeast than it is in mammals. At least nineteen different RNAs are 'cappable' in vitro with the enzyme guanylyl transferase. This enzyme distinguishes the 5' ends of primary transcripts from those generated

by processing. Thus these nineteen RNAs presumably represent the products of de novo transcription-initiation at distinct promoters (Edwards et al 1983). To date, fourteen of these primary transcripts have been localised on the genetic map (Christianson et al 1982, Christianson et al 1983).

Comparison of the DNA sequences at the sites of initiation of transcription reveals a conserved nonanucleotide sequence 5'$\binom{T}{A}$ATAAGTA 3' (reviewed by Tabak et al 1983); this sequence also precedes the rRNA genes of the fungus Kluyveromyces lactis (Osinga et al 1982) and may thus represent part of a promoter. In all cases examined transcription is initiated at the last A residue of the conserved sequence.

Several genes lack this sequence motif and it seems likely that they form part of larger transcriptional units and their mature mRNAs are subsequently generated by processing. For example no putative promoter is found between the 5' end of the COB gene and a $tRNA^{Glu}$ gene 1050 bp upstream (Christianson et al 1983). The gene clusters thought to be under the control of one 'promoter' are

5' 21S rRNA .....$tRNA^{Thr}$ 3'	Levens et al 1981
5' $tRNA^{Phe}$..$tRNA^{Val}$....COIII 3'	Thalenfeld et al 1983
5' $tRNA_f^{met}$...tRNA synthesis locus...$tRNA^{Pro}$	Miller et al 1983
5' COI...ATPase 8...ATPase 6... URF 3'	Edwards et al 1983
5' $tRNA^{Glu}$....COB 3'	Christianson et al 1983
5' ATPase 9... $tRNA^{Ser}$... Var-1 3'	Martin et al 1982, Edwards et al 1983

While the position and evolutionary conservation of this sequence suggest that it may function as a promoter in fungal mitochondria, there is evidence that it alone is insufficient for transcription

initiation and other factors must be involved. The evidence is:

1) Edwards et al (1982) have used an in vitro transcription system using purified S. cerevisiae mt RNA polymerase to analyse 'run-off' transcripts from mt DNA templates. While S. cerevisiae templates are correctly transcribed in vitro, K. lactis templates, which contain perfectly conserved copies of the sequence, are not (Tabak et al 1983). Hence the presence of a 'promoter' in front of a gene is not sufficient to specify correct initiation.

2) Controlled mutagenesis (Zoller and Smith 1982) shows that the 'promoter' can tolerate drastic sequence changes and still direct correct transcription initiation (Tabak et al 1983).

3) Two 'promoter' sequences are found upstream of ATPase 9 gene and yet only the upstream 'one' directs high level transcription (Edwards et al 1983, Zassenhaus et al 1984).

It is also clear that tRNAs, mRNAs and rRNAs are under the control of one 'promoter'. The fact that the same purified polymerase preparation is capable of correct initiation at sites serving to transcribe mixed collections of genes strongly suggests that a single mt RNA polymerase is used to transcribe all three types of mitochondrial gene. Transcription in other fungal species has not been so extensively characterised as the yeast and fungal mitochondria. However, a sequence showing 70% homology to the yeast nonanucleotide putative promoter sequence has been found preceding the Neurospora crassa COI gene (de Vries et al 1983). Tandemly repeated PstI sites (PstI palindromes) are thought to play a role in the processing of polycistronic transcripts in N. crassa (Yin et al 1980) perhaps by providing a recognition site for a processing enzyme. Thus for example the MAL gene (ATPase proteolipid) is thought to be cotrans-

cribed with the upstream ATPase 6 gene and the 5' end of the processed MAL transcript maps precisely to a PstI palindrome (de Vries et al 1983)

In yeast, a conserved dodecamer sequence 5' AAUAAUAUUCUU 3' has been implicated in processing (distinct from splicing) of polycistronic mRNAs (Osinga et al 1984a). Processing of yeast mt tRNAs at their 5' ends is dependent on the action of an enzyme which requires a specific RNA component encoded by the mitochondrial tRNA synthesis locus (Miller et al 1983). E. coli tRNAs are processed in a similar way (Hopper et al 1982) and indeed yeast pre-tRNAs can be accurately processed by E. coli RNase P (Miller et al 1983).

The splicing (as opposed to processing) of the transcripts of split mt genes has been studied in particular detail. Excision of introns is dependent on one or more of the following: self encoded proteins (maturases), nuclear encoded products, critical sequences within the intron but removed from the splice site and formation of complex secondary structures in the unspliced transcript. This subject has been reviewed extensively by Grivell et al (1983) and is only briefly treated here. The splicing pathway of the 'long' form of the yeast COB gene containing five introns is shown in Figure 1.3. No maturase has yet been isolated in active form and their role in splicing is as yet speculative. Until recently there was no evidence that the open reading frames present in some introns postulated to encode part of a maturase were actually translated (although there was very good genetic and biochemical evidence suggesting the existence of a diffusible protein product). However using antibodies raised against synthetic oligopeptides (Guiso et al

Fig. 1.3 Splicing pathway of the 'long form' of the COB gene in S. cerevisiae

The splicing pathway of the 'long form' of the S. cerevisiae COB gene (strains 777-3A and KL 14-4A) is as follows:

1) Intron 1 is excised from the pre-mRNA by a nuclear encoded enzyme. This generates a 10S circular RNA.

2) Exon 1 is joined to exon 2, and together with an open reading frame in intron 2 form the mRNA for a maturase enzyme.

3) Maturase splices intron 2, destroying its own mRNA.

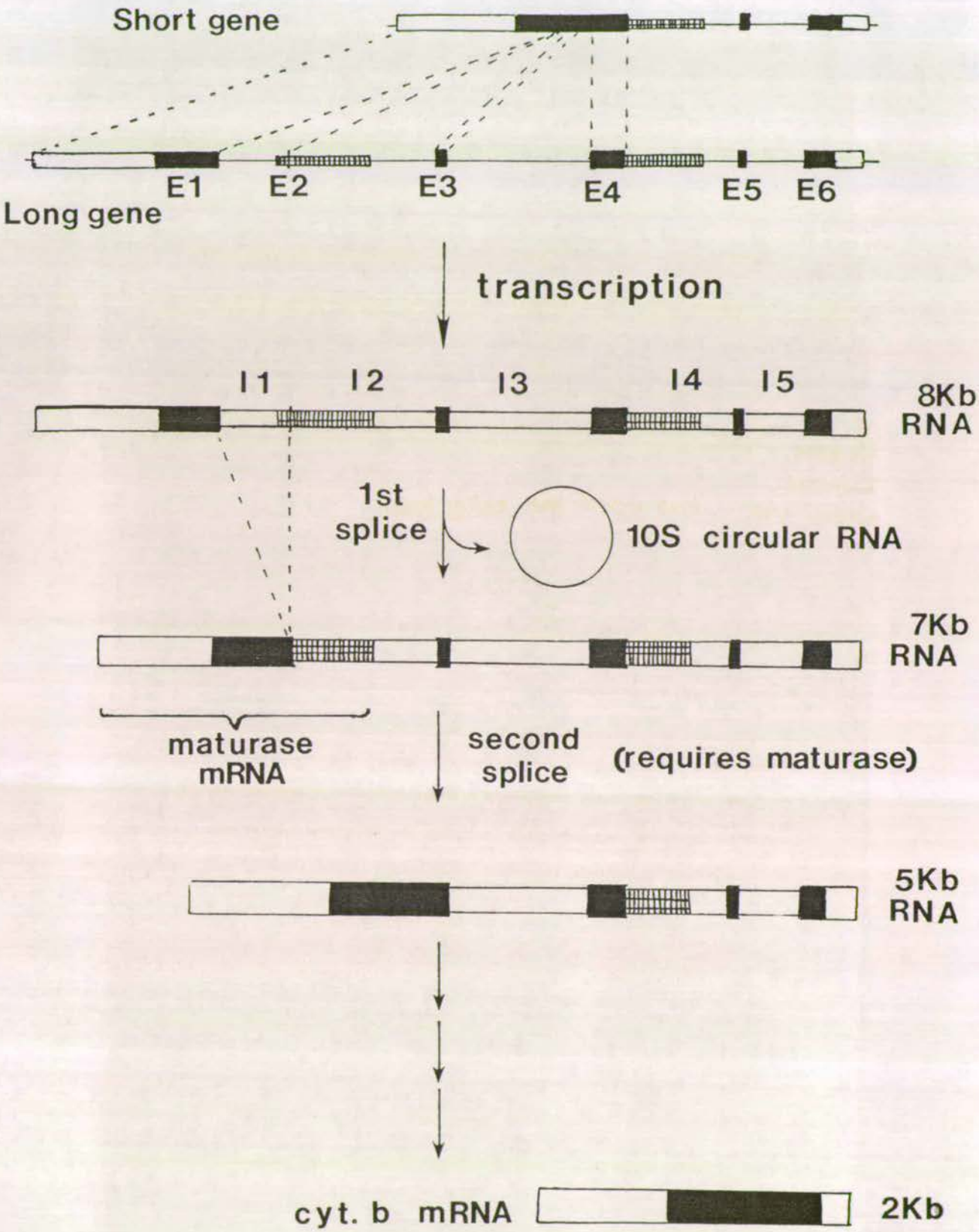
4) Intron 3 is spliced by a nuclear enzyme.

5) Exons 1, 2, 3 and 4 form mRNA for a maturase which splices intron 4.

6) Intron 5 is spliced by a nuclear enzyme.

E = exon
I = intron
Hatched area = intron encoded open reading frame

Fig. 1.3


Short gene
E1
E2
E3
E4
E5
E6
Long gene
transcription
I1
I2
I3
I4
I5
8Kb RNA
1st splice
10S circular RNA
7Kb RNA
maturase mRNA
second splice
(requires maturase)
5Kb RNA
cyt. b mRNA
2Kb

1984) and fusion proteins (Jacq et al 1984), the products of introns 2 and 4 of the yeast COB gene have been identified as proteins of 42,000 and 27,000 respectively in splicing deficient mutants which accordingly overproduce the maturase. In wild type cells, the maturase is not detected. Interestingly the maturase encoded by COB intron 4 is not only important in the splicing of COB transcripts, but also appears to be involved in the expression of the COI gene, whereby it excises intron 4 of COI from the unspliced transcript (Church et al 1979, Dhawle et al 1981). Mutations in intron 4 of COB obviously affect correct processing of COI, but can be suppressed by both nuclear (NAM 2-1), and mitochondrial (mim 2-1) mutations (Groudinskyet al 1981). Both mim 2-1 and NAM 2-1 mutations are thought to activate a previously silent maturase in COI intron 4. The former causes an amino acid change in the COI intron 4 (Dujardin et al 1982); the mode of action of NAM 2-1 is not known (Dujardin et al 1983).

The identification of two nuclear mutants defective in processing of mt transcripts points to the further involvement of nuclear genes in yeast mt gene expression. Mutations in the nuclear CBP-1 gene introduce an instability to COB transcripts due to incorrect 5' processing (Dieckmann et al 1984a, 1984b), and a defective nuclear encoded MSS-51 product prevents proper splicing of COI introns (Faye and Simon 1983, Simon and Faye 1984).

1.6.4 Plant mitochondrial RNAs

Transcription of the higher plant mt genome has barely been investigated, and this area forms the basis of the work described in Chapters 3, 4 and 5.

1.6.5 RNA synthesis in isolated mitochondria

Isolated yeast mitochondria incorporate ^{32}P UTP into RNA which hybridises to mt DNA. Incorporation of nucleotides into RNA proceeds linearly for over one hour, and the products are insensitive to ribonuclease, strongly suggesting synthesis has taken place in intact mitochondria (Groot et al 1981, Boerner et al 1981, Newman and Martin 1982).

The labelling kinetics of the yeast 21S and 15S rRNAs have been studied using isolated mitochondria. The 21S pre rRNA (which is split by a 1100b intron (Groot et al (1981)) and undergoes processing at the 3' end whereby 1 Kb of 3' extension is removed (Marten et al 1980), undergoes complete processing and splicing during a chase (Boerner et al 1981). The 15.5S precursor rRNA could not however be chased into mature 15S rRNA (Boerner et al 1981). This suggests that some factors normally supplied from the cell sap are lacking in isolated mitochondria and thus prevent proper maturation.

1.7 TRANSLATION OF MITOCHONDRIAL RNAs

The mitochondrial translation system is distinct from that of the cytoplasm (Kuntzel and Noll 1967) and shares a number of features with the bacterial system (reviewed in chapter 3). The majority of components are encoded by the nucleus, the exceptions being rRNAs, tRNAs and a ribosomal protein in fungi and plants. The mitochondrial ribosomes from a variety of organisms differ greatly in their sedimentation coefficients (Table 1.3).

Table 1.3 Sedimentation coefficients (S) of mt ribosomes and rRNAs

Organism	Ribosomes (S)	rRNAs (S)	Reference
Trypanosomes	-	12S 9S	
Humans	55 - 60	16S 12S	Anderson et al 1981
Yeast	73 - 74	21S 15S	Tzagoloff 1982
Plants	77 - 78	26S 18S 5S	Reviewed in Leaver and Gray 1981
Schizosaccharomyces pombe	-	19S 14S	Lang et al 1983

Mammalian and fungal mitochondria encode only a small number of tRNAs (22 and 25 respectively, Anderson et al 1981, Borst and Grivell 1981). According to "wobble rules", 32 tRNAs are required to decode the 61 amino acid encoding triplets. It appears that the tRNAs of these mitochondria have an expanded codon recognition pattern and a "2 out of 3" pairing process operates between codon and anti-codon (Attardi 1981). Some mt tRNAs show unusual structure eg human mt $tRNA^{Ser}$ which lacks one of the 'universal' arms (de Bruijn et al 1980).

Little is known about mitochondrial translation initiation. Since mammalian mt mRNAs do not carry 5' non-translated regions, translation must start at or very near the 5' end (Montoya et al 1981), and it is not known how specificity of translation initiation is conferred. Putative ribosome binding sites (originally described for eubacterial and chloroplast mRNAs, Shine and Dalgarno 1974, Steitz and Jakes 1975, Whitfeld and Bottomley 1983) have been identified in fungal and plant mt RNAs which do carry 5' untranslated regions. Li et al (1982) have suggested that in yeast mitochondria the sequence 5' AAAUUCUAUA 3' found at the 3' end of the 15S rRNA could base pair

with a 4 - 10 nucleotide complementary sequence located 8 - 116 nucleotides upstream from all S. cerevisiae protein coding genes. A more convincing putative ribosome binding site has been identified in plant mitochondria (Dawson et al 1984) involving the octanucleotide sequence 5' UGAAUCCU 3' at the 3' end of maize 18S rRNA (Chao et al 1983). Analysis of a variety of plant mt protein coding genes shows many are preceded 13 - 18 nucleotides 5' to the AUG codon by a tetranucleotide displaying 3/4 or 4/4 complementarity with the 18S rRNA sequence 5' UGAA. Additional homology to the remainder of the octanucleotide also occurs in this region (Dawson et al 1984).

The biochemistry of mitochondrial protein synthesis has not been studied in any detail. However it is clear that the nucleus plays an important role since it encodes components of the translation system. For example, mammalian mitochondrial ribosomes contain at least 85 proteins (O'Brien and Matthews 1976), none of which is thought to be encoded on the mt genome. Recently the nucleus has also been shown to encode proteins which specifically stimulate translation of particular mt mRNAs. Thus in S. cerevisiae mutations of the nuclear gene PET 494-1 prevent translation of COIII mRNAs; this mutation can be suppressed in petites where the COIII gene is fused to the 5' ends of genes which are not under the control of the PET 494-1 product (Müller et al 1984).

1.8 MITOCHONDRIAL MUTATIONS AND CYTOPLASMIC MALE STERILITY

1.8.1 Mitochondrial mutations; general features

Mutations in mitochondrial DNA which disrupt mitochondrial function have greatly expedited the identification of mt genes in S. cerevisiae.

These mutations are not lethal because yeast can survive by anaerobic fermentation. In other organisms complete deficiency in the synthesis of any mitochondrial translation product is lethal and so the number of mitochondrial mutants isolated tends to be small. Many mitochondrial mutants have been isolated and characterised from S. cerevisiae, and it is the number and diversity of these mutants that has allowed detailed genetic maps of the yeast mt genome to be constructed. Rho$^-$ mutants (petites) occur spontaneously at very high frequency and result from the deletion of large sequences of mt DNA and the tandem amplification of the remaining segments. Many point mutations (mit$^-$ syn$^-$ and mim reviewed in Bernardi 1983) have now been mapped within both coding sequences and introns. Syn$^-$ mutations map in the genes of the protein synthetic machinery and lead to complete or partial deficiency in mitochondrial protein synthesis. Mit$^-$ mutations map within protein coding genes and thus affect respiration. Mim mutations restore respiratory capacity to other mitochondrial mutants but do not themselves confer a mutant phenotype eg mim-2-1 described in section 1.6.3.

Excisions and amplification of mt DNA similar to petites have also been described in other fungi, notably senescent cultures of Podospora anserina (Wright and Cummings 1983), the poky and stopper mutants of Neurospora crassa (de Vries et al 1981) and the ragged mutants of Aspergillus amstelodami (Lazarus and Kuntzel 1981).

1.8.2 Cytoplasmic male sterility in plants

Cytoplasmic male sterility (cms) is a maternally inherited trait which prevents the formation of functional pollen but does not affect female fertility (Duvick 1965, Edwardson 1970). This phenotype has in the

past been commercially useful in production of hybrid corn by preventing self pollination of the seed parent. There is considerable evidence (reviewed in section 1.8.2.1 and by Leaver and Gray 1982) that the cms trait is carried on the mt DNA.

In maize there are three types of cms, cms (T, S and C) which differ in the specific nuclear genes required to restore fertility (restorer, Rf genes, Beckett 1971). The cms mutations are conditionally lethal, since they only cause dysfunction when the plant is placed under certain types of stress for example during anthesis. Maize plants carrying the T cytoplasm are also susceptible to the T toxin produced by the fungus Dreschlera maydis Race T (Ullstrup 1972). This toxin causes uncoupling of oxidative phosphorylation and leakage of NAD^+ from mitochondria, leading to cell death (Matthews et al 1979). The toxin is therefore another form of stress but one which causes the cms mutation to be manifested throughout the whole plant.

1.8.2.1 Evidence for the association of cms with a mitochondrial mutation

1. Mitochondrial DNAs isolated from male fertile (N) or cms-C S and T plants can be distinguished on the basis of their restriction enzyme profile (Pring and Levings 1978) whereas the chloroplast DNAs of the same cytoplasms are almost identical.

2. Isolated mitochondria from N C S and T plants synthesise a common spectrum of 18 - 20 polypeptides, each male sterile cytoplasm being characterised by the synthesis of additional polypeptide(s). Cms T and C mitochondria synthesise polypeptides of M_r 13,000 and 17,500 respectively which replace products of M_r 21,000 and 15,000 in N. Cms S mitochondria synthesise eight additional polypeptides of M_r 42 - 85,000 (Forde et al 1978, Forde and Leaver 1980).

Forde and Leaver (1980) suggest that the 13,000 M_r polypeptide synthesised by cms T mitochondria could act as a binding site for T toxin and for some anther-specific substance. Binding of these substances could lead to changes in mitochondrial membrane permeability and leakage of NAD^+. The ability of the restorer genes to suppress synthesis of the 13,000 M_r polypeptide may be relevant to this model, although Walton (1983) suggests absence of the 21,000 M_r polypeptide rather than presence of 13,000 M_r polypeptide is more closely correlated with toxin sensitivity.

3. The cms T genotype and susceptibility to D. maydis T toxin are inseparable, suggesting that both traits are different manifestations of the same mutation (Brettel et al 1980, Gengenbach et al 1981). Thus the specific susceptibility of mitochondria from T cytoplasm to T-toxin while chloroplasts from the same cytoplasm (and also mitochondria from N cytoplasm) are unaffected implies the mutation resides in the mitochondrion.

4. Ultrastructural studies during pollen formation show mitochondrial degeneration in the anthers is the first sign of abnormality in cms T plants. This degeneration coincides with the end of a period of rapid mitochondrial division, again reflecting the conditional nature of the mutation (Warmke and Lee 1977).

1.8.2.2 Plasmid-like DNA molecules in plant mitochondria

Plasmid-like episomal DNA molecules have been found in a variety of plant mitochondria eg. sugarbeet (Powling 1981), Brassica species (Palmer et al 1983), sorghum (Dixon and Leaver 1982, Pring et al 1982), and maize (Pring et al 1977). A brief discussion of two of the episomal DNAs of maize is included here because of their possible association with cytoplasmic male sterility.

Cms S mitochondria contain equimolar amounts of two linear molecules of 6.4 and 5.4 Kb designated S1 and S2 respectively (Pring et al 1977). These 'plasmids' have the following properties:

1) Cms S mitochondria normally contain equal amounts of S1 and S2 except in certain 'Vg' lines (a type of cms S cytoplasm) where the relative proportions of S1 and S2 appears to be under nuclear control (Levings et al 1980). S1 and S2 are present in a fivefold molar excess over 'main band' mt DNA (Thompson et al 1980).

2) S1 and S2 share homology of approximately 1,500 bp at one end. Both have terminal inverted repeats of approximately 200 bp (Levings and Sederoff 1983).

3) S1 and S2 are covalently attached to protein molecules at their termini (Kemble and Thompson 1982). Levings (pers. commun.) suggests that these proteins may be involved in the replication of S1 and S2 which could occur in a similar fashion to that of Adenovirus DNA (Rekosh et al 1977).

4) Sequence analysis of S2 (Levings and Sederoff 1983) shows the presence of two open reading frames on opposite strands of 3613 and 1017 bp. S1 contains three open reading frames, the largest of which is 2787 bp (S. Levings pers. commun.). These are clearly long enough to encode the high molecular weight proteins characteristic of cms S cytoplasm, but in the absence of a cell free transcription/translation system for mt DNA (Chapter 3) this relationship cannot be proven other than by microsequencing of the HMW proteins and comparison of the protein and DNA sequences.

5) Cms S strains are unstable and revert to male fertility spontaneously (Laughnan and Gabay 1973). These reversions can be at the nuclear or cytoplasmic level. The former are functionally equivalent

to the generation of new Rf loci, although individual nuclear revertants map at several different points in the nuclear genome and even on different chromosomes (Laughnan and Gabay 1975). Cytoplasmic reversions to fertility are associated with the loss of free S1 and and S2 from the mitochondria and the appearance of novel 'main band' mt DNA restriction fragments showing homology to S1 and S2 (Levings et al 1980). The correlation of integration of S1 and S2 into mt DNA with reversion to fertility suggests they could be acting as mobile male fertility elements (Laughnan and Gabay 1975a), although amplification of existing sequences cannot be discounted.

6) Sequences homologous to S1 but not S2 have been found in the nuclear DNA of N, cms S, SRf and nuclear revertant lines. The restriction fragments hybridising to S1 are identical in each case showing the S1 is not acting as a mobile 'fertility element' in nuclear reversion (Kemble et al 1983).

7) Sequences homologous to portions S1 and S2 have been found in 'main band' mt DNA of N mitochondria (Thompson et al 1980). The integrated copies lack the 1,500 bp homologous region (McNay et al 1983), and they are flanked at one end by the 3 Kb repeat sequence thought to be involved in intramolecular recombination (Lonsdale et al 1983 and section 1.3.4). S1 and S2 specific sequences hybridise only weakly to 'main band' mt DNA from C S and T cytoplasms. Such faint hybridisation may be due to

1) short stretches of homology in main band mt DNA from these cytoplasms or

2) a variant sub population of mt genomes carrying integrated copies of S1 and S2.

Other plasmid-like DNAs have been found in various maize lines. Some races of S. American corn contain two linear episomes R1 and R2 which are not associated with male sterility but show considerable homology to S1 and S2. R2 is probably identical to S2 while R1 is similar to S1 but is approximately 1 Kb longer (Weissinger et al 1981, 1983). Levings (1983) has suggested S1 might have arisen by recombination between R1 and R2, thus accounting for the origin of the 1.5 Kb shared sequence. Other linear and closed circular episomal DNAs have been found in N C S and T mitochondria (see for example Kemble and Bedbrook 1980, Kemble et al 1983a).

1.9 MIGRATORY DNA

One of the most unexpected findings in the last few years has been the discovery that organelle and nuclear genomes are not maintained discretely in their respective organelles, but rather share common sequences, dubbed 'promiscuous DNA' by Ellis (1982). Maize mitochondria have been shown to contain two stretches of DNA homologous to chloroplast DNA. One contains a highly conserved form of the gene for the large subunit of RuBPCase, and the other is a 12 Kb stretch containing the 16S rRNA and two to three tRNA genes (Lonsdale et al 1983a, Stern and Lonsdale 1982). Stern and Palmer have shown that chloroplast DNA sequences are found in the mt genomes of a variety of angiosperms including the chloroplast β ATPase which is found in mung bean mt DNA. The identification of a partial copy of photogene 32 (psbA) in S1 DNA (S. Levings pers. commun.) means this chloroplast sequence is present in both main band and episomal mt DNA, and also possibly therefore integrated into the nucleus in N, cms S SRf and nuclear revertants (section 1.9.2.2). Interorganelle

movement is not restricted to chloroplasts and mitochondria. Nuclear copies of mitchondrial sequences have been found in a variety of organisms for example yeast (Farelly and Butow 1983), locust (Gellissen et al 1983), rat (Hadler et al 1983) and Podospora anserina (Wright and Cummings 1983). Chloroplast sequences have been found in the nucleus of spinach (Timmis and Scott 1983). In most cases the DNA sequences integrated into the nuclear genome represent partial copies of genes which are often rearranged. However in sea urchin, a complete copy of COI has become integrated and TGA codons appear to have been specifically converted to TGG (Jacobs et al 1983). Thus if transcribed the nuclear copy could be translated to yield a complete COI protein; however there is no evidence that it is transcribed in the nucleus. This may represent the early transfer of a mitochondrial gene to the nucleus, later stages being characterised by the mitochondrial copy becoming non-functional. This premise is supported by the finding that ATPase subunit 9 gene is encoded by the nuclear genomes of N. crassa (van den Boogart et al 1982) and A. nidulans (Scazzocchio et al 1983) and yet the mitochondria retain silent non-expressed copies of these genes. The way in which DNA is transferred between organelles is not known. Ellis (1982) suggests three mechanisms: breakdown of organelles followed by uptake of DNA into novel locations; fusion of organelles followed by recombination; or vector mediated transfer. It seems likely that these different potential transfer mechanisms may be used by different 'promiscuous' sequences in the exchange of DNA sequences between organelles of different organisms.

1.10 SCOPE OF THIS THESIS

It is apparent from the review of the literature presented in this chapter that while the mitochondrial genes and their expression have been well characterised in mammals and fungi, an equivalent understanding of plant mitochondrial genes has not yet been reached. The work described in this thesis was undertaken to partly redress that balance. Three particular aspects were studied in detail:

1) To identify coding regions of the higher plant mitochondrial genome, attempts were made to develop a cell free system which would transcribe and translate mt DNA with fidelity. Such a system would greatly expedite our understanding of the coding capacity of the higher plant mt genome, which is likely to contain genes not found in the mt DNA of other organisms, and which therefore could not be easily identified. A cell free system capable of translating mt RNA would also allow studies on the expression of the plant mt genome, allowing for example quantitative measurements of translatable RNAs during mitochondrial biogenesis.

2) To study the way in which higher plant mt genes are expressed, the transcripts of four protein coding genes (COI, COII, COB and F_1 ATPase α subunit) and of the episomal DNAs S1 and S2 were preliminarily characterised. In the absence of a cell free transcription translation system, identification of RNAs hybridising to a gene specific DNA probe is good evidence that a particular gene is actually expressed in the mitochondrion.

3) A preliminary study on the expression of mt genes during biogenesis of mitochondria was undertaken. During germination and early development of Vicia faba seedlings, there is a great differential in the ability of mitochondria to incorporate ^{35}S methionine into

protein. There are also differences in the spectrum of polypeptides synthesised. Preliminary investigations as to whether these differences are due to a requirement for mitochondrial mRNA synthesis were undertaken.

CHAPTER 2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Seed

Maize seed was provided by Pioneer Hibred International Des Moines, Iowa, U.S.A. Seeds with the nuclear background 3541 and cytoplasmic genotype N (normal) were used for most of the studies in this thesis. Seed with nuclear background WF9 with N (normal) and C, S and T (male sterile) cytoplasmic genotypes was used in experiments as stated. Vicia faba (Faba bean) seed was provided by the National Seed Development Organisation.

2.1.2 Chemical reagents

All reagents other than those listed below were obtained from British Drug Houses (BDH Ltd., Poole) or from the Sigma Chemical Co. (Poole) and were of Analar or Sigma grade. Other reagents used were:

Acrylamide and NN'Methylenebisacrylamide: BDH 'Electran' grade

Agarose: Miles Laboratories

'Sea Kem' Agarose: FMC Corporation, Rockland, U.S.A.

Ampicillin ("Penbritin"): Beecham Research Laboratories

Butyl PBD: Intertechnique Ltd.

Caesium chloride: Fisons Scientific Apparatus

Coomassie Brilliant Blue R: Raymond A. Lamb Ltd.

Deoxynucleotide triphosphates and yeast tRNA: Boehringer Mannheim GmBH

Dideoxynucleotide triphosphates: P - L Biochemicals Inc.

Herring Sperm DNA: Serva Feinbiochemica, Heidelberg

PVP: L. Light & Co. Ltd.

Sephadex: Pharmacia fine chemicals

Triton X - 100: Hopkin and Williams

2.1.3 Enzymes

Restriction enzymes were obtained from the following suppliers:

Amersham International plc, Bethesda Research Laboratories (U.K.) Ltd.

Boehringer Mannheim GmBH, P & S Biochemicals Ltd.

Calf intestinal phosphatase (CIP) was supplied as an ammonium sulphate suspension from B.R.L. To prepare for use it was passed through a Sephadex G75 column equilibrated with 20mM Tris Cl pH 7.4 and 100mM KCl. Fractions containing CIP were identified by their ability to prevent ligation of restriction-enzyme-digested DNA.

Active fractions were pooled and glycerol added to 50% (v/v)

DNase I (RNase free): Worthington Biochemical Corporation

DNA polymerase I (*E. coli*): New England Biolabs

DNA polymerase I, Klenow fragment (*E. coli*): Boehringer or BRL.

T4 DNA ligase, T4 polynucleotide kinase and lysozyme: Boehringer

All enzymes were stored at -20°C according to the manufacturers' instructions.

2.1.4 Radioisotopes

The following radioisotopes were purchased from Amersham International plc:

α ^{32}P dCTP (triethyl ammonium salt in stabilised aqueous solution) at 400 Ci/mmol, 10 μCi/μl

^{35}S-L-methionine (in 20mM potassium acetate, 0.1% (w/v) 2-mercaptoethanol, 1mM DTT), 1000 Ci/mmol, 5 μCi/μl

γ ^{32}P dATP at 4000 Ci/mmol and 10 μCi/μl was purchased from New England Nuclear

2.1.5 Stock Buffers and Media

TE80: 10mM Tris, 1mM EDTA pH adjusted to 8.0 with 1M HCl

10 x TBE: 0.89M Tris, 0.89M Boric acid, 0.02M EDTA pH 8.3

Phenol: redistilled under degassed water and equilibrated with TE80 until pH 8.0

20 x SSC: 3M NaCl, 0.3M Sodium Citrate pH adjusted to 7.0 with 1M HCl

Minimal agar: 1.5% (w/v) Difco bacto agar, 0.2% (w/v) $(NH_4)_2SO_4$, 1.4% (w/v) KH_2PO_4, 0.6% (w/v) KH_2PO_4, 0.1% (w/v) Sodium Citrate, 0.02% (w/v) $MgSO_4$, 0.2% (w/v) glucose, 2.5×10^{-4}% (w/v) thiamine hydrochloride.

BBL top agar: 1% (w/v) Difco bacto agar, 0.65% (w/v) Baltimore Biological Laboratories trypticase, 0.5% (w/v) NaCl

L. Broth (Luria and Delbruck, 1943): 1% (w/v) Difco Bacto Tryptone, 0.5% (w/v) Difco Bacto yeast extract, 0.5% (w/v) NaCl, pH adjusted to 7.2 with 5M NaOH.

L agar: 1.5% (w/v) Difco agar, 1% (w/v)Difco Bacto tryptone, 0.5% (w/v) Difco Bacto yeast extract, 1% (w/v) NaCl pH adjusted to 7.2 with 5M NaOH.

All buffers and media were either autoclaved at 15 p.s.i. for 20 minutes or prepared from individually sterilised components.

2.1.6 Miscellaneous

Small scale reactions were carried out in 1.5ml polypropylene tubes (Sarstedt) which were centrifuged as required in a 12 or 40 place Eppendorf centrifuge at a fixed force of $12,000 \times g_{max}$. Larger volumes were handled in 15ml and 30ml Covex tubes or 250ml polycarbonate bottles. These were centrifuged as necessary in a Sorvall RC5B centrifuge using SS-34 and GSA fixed angle rotors,

and an HB4 swing-out rotor, up to 40,000xg_{ave}

Ultracentrifugation up to 100,000xg_{ave} was carried out either in a Sorvall OTD50B centrifuge using a fixed angle TFT.65.13 rotor or a Spinco L ultracentrifuge using a fixed angle Type 40 rotor. All tubes and bottles were sterilised before use by autoclaving (1.5 ml polypropylene tubes), baking at 120°C overnight (Corex tubes), or by immersion in 1% (w/v) sodium hypochlorite for 30 minutes.

Seeds were grown on 12 ply cellulose wadding (Robinson and Sons Ltd.)

2.2 METHODS

2.2.1 Isolation of maize mitochondrial RNA

2.2.1.1 Buffers

GrindingMedium: 0.4M mannitol, 1mM EGTA, 25mM MOPS, 0.1% (w/v) BSA, 8mM cysteineHCl pH adjusted to 7.8 with 5M KOH

Wash Medium: 0.4M mannitol, 1mM EGTA, 5mM MOPS, 0.1% (w/v) BSA pH adjusted to 7.5 with 0.5M KOH

Buffered sucrose stock solution: 2M sucrose, 10mM Tricine, 1mM EGTA pH adjusted to 7.2 with 5M KOH. The stock solution was diluted for gradients with appropriate volumes of 10mM tricine, 1mM EGTA pH 7.2

Dilution Medium: 0.2M mannitol, 10mM Tricine, 1mM EGTA adjusted to pH 7.2 with 0.5M KOH

RNA extraction buffer: 0.5% (w/v) NDS, 50mM Tris, 50mM NaCl pH adjusted to 7.4 with 1M HCl.

1 x DNase buffer: 50mM MES, 2.5mM Magnesium acetate pH adjusted to 7.0 with 10M NaOH.

2:2:1:2 Growth of plant material

Seeds were surface sterilised for 15 minutes in sodium hypochlorite [1% (w/v)] and imbibed overnight in cold running water. The seeds were planted on sterile cellulose wadding moistened with sterile distilled water, and allowed to germinate in darkness for 4 or 5 days at 30°C. Each tray (45 x 27 cm) contained 50 - 100g seed.

2.2.1.3 Isolation of mitochondria

Mitochondria were isolated according to Leaver et al (1983). All operations after harvesting were carried out at 4°C. Maize coleoptiles were harvested into ice-cold water, cut into 1cm sections and washed three times with ice-cold sterile water. After the addition of grinding medium (2ml /g tissue), the tissue was homogenised to a creamy consistency in a pestle and mortar. The homogenate was filtered through 4 layers of muslin, 2 layers of milk filter (Blow, "Maxa") and centrifuged at $1000xg_{ave}$ for 5 minutes in the Sorvall GSA rotor. The supernatant was transferred to clean bottles and re-centrifuged at $14,000xg_{ave}$ for 15 minutes. The mitochondrial pellet from approx. 50g tissue was resuspended in 20mls of wash medium and subjected to one more round of differential centrifugation. The enriched mitochondrial pellets were resuspended in wash medium (2ml / 50g tissue) using a small homogeniser and further purified by centrifugation through a continuous sucrose gradient. The sucrose gradients were prepared by layering over 3ml 2M sucrose, 7.5ml each of 1.45, 1.2 and 0.9M and 4ml of 0.6M sucrose in a 36ml polyallomer tube, and allowed to diffuse overnight at 4°C. The equivalent of approximately 50g of starting tissue was layered onto each gradient. The mitochondrial suspension was overlaid with 4ml of wash medium (diluted 1:1 with water) and centrifuged in the SS-90 vertical rotor

at 21,000xg_{ave} for 45 minutes. The mitochondrial band at 1.25M sucrose was recovered from the gradients and diluted slowly by the addition of 1.5 volumes of dilution buffer. The mitochondria were pelleted by centrifugation at 10,000xg_{ave} for 15 minutes. This pellet of purified mitochondria was used directly to prepare mt RNA.

2.2.1.4 Isolation of mitochondrial RNA

Mitochondrial RNA was extracted by the method of Koller et al (1982). The mitochondrial pellet was gently resuspended in 1 - 2mls of RNA extraction buffer per 100g starting material, causing the mitochondria to lyse. Mitochondrial nucleic acids (RNA + DNA) were extracted by the addition of an equal volume of phenol. The aqueous and phenol phases were separated by centrifugation and the phenol back-extracted with an equal volume of RNA extraction buffer. The combined aqueous phases were made 6% (w/v) with respect to PAS and phenol extracted twice more. The nucleic acids were precipitated overnight at +2°C by the addition of sodium acetate to 0.3M and 2.5 volumes of ethanol. The precipitated nucleic acids were pelleted by centrifugation at 14,000xg_{ave} for 30 minutes, washed 3 times in 70% (v/v) ethanol and resuspended in 50-100µl of sterile water. The yield of mt nucleic acid was estimated by measuring A_{260}, assuming that 1 O.D. unit at 260nm is equivalent to 40µg/ml nucleic acid (Maniatis et al 1982). 1g of tissue generally yielded 1 - 2µg mt nucleic acid. An indication of the purity of the mt nucleic acid was obtained by calculating the ratio $A_{260}:A_{280}$ which was always approximately 2:1.

2.2.1.5 Removal of DNA from mt nucleic acid

DNA was removed from the total nucleic acids by the method of Silverthorne and Ellis (1980). An equal volume of 2 x DNase buffer

was added to the mt nucleic acid and DNase (in 1 x DNase buffer) added to a final concentration of 20µg/ml. Digestion was allowed to continue on ice for 20 minutes, after which time the remaining RNA was made 200mM with respect to NaCl and the reaction terminated by phenol extraction. The aqueous phase was ethanol precipitated for 30 minutes at -80°C, washed twice with 70% (v/v) ethanol, desiccated and resuspended in sterile water to a final concentration of 1µg/µl. Solid NaCl was added to the mt RNA to a final concentration of 2.5M. RNA was selectively precipitated overnight at +2°C, pelleted by centrifugation at 12,000xg_{max} for 30 minutes., washed twice with 2.5M NaCl and twice with 70% (v/v) ethanol. The RNA was then desiccated and resuspended in water to a final concentration of 5µg/ul. RNA was stored at -20°C.

2.2.2 Extraction of nucleic acid from plant tissues and E. coli

2.2.2.1 Buffers

Nucleic acid extraction buffer: 6% (w/v) PAS, 1% (w/v) TNS, 100mM Tris pH adjusted to 8.5 with 1M HCl. Prepared just before use.

2.2.2.2 Purification of total nucleic acid from plant tissue

DNA and RNA was extracted from plant tissue using detergent and phenol (Parish and Kirby 1966). 5-day old maize shoots (5 - 10g) or 10 - 20 cotyledons of Vicia faba were quick-frozen in liquid N_2. The tissue was homogenised immediately in 2 volumes of nucleic acid extraction buffer using a pestle and mortar or an MSE homogeniser at half speed until a creamy consistency was obtained. The aqueous phase was extracted by the addition of an equal volume of phenol and centrifuged for 5 minutes at 1300xg_{ave} in an MSE Mistral 4L centrifuge to resolve the phases. The aqueous phase was phenol extracted twice

more or until no more proteinaceous matter was seen at the interface. Total nucleic acids were precipitated overnight at -20°C after the addition of sodium acetate to 0.3M and 2.5 vols of ethanol. The nucleic acids were collected by centrifugation 1,300xg_{ave}, 30 mins., washed three times in 70% (v/v) ethanol and resuspended in sterile water. Typical yields were 250µg/g tissue.

2.2.2.3 Extraction of total nucleic acids from E. coli

A single colony of E. coli CSH 73 (see section 2.2.3.2.1) was used to inoculate a 5ml 'overnight' culture which was grown to stationary phase. This was used to inoculate 4 x 100ml cultures and bacteria grown and harvested exactly as described in section 2.2.3.2.1. The final bacterial pellet was quick frozen in liquid N_2 and stored at -80°C until required. The frozen bacterial pellets were homogenised in 10ml nucleic acid extraction buffer per g material using a Polytron homogeniser (Northern Media Supply Ltd., Hull) at full speed for 10 seconds. Thereafter the nucleic acid extraction was identical to that described in section 2.2.2.2.

2.2.3 In vitro transcription and translation of nucleic acids in an E. coli S-30 extract

2.2.3.1 Buffers

Buffer A: 10mM Tris, 10mM Magnesium acetate, 60mM Ammonium acetate, 6mM 2-mercaptoethanol pH adjusted to 8.2 with glacial acetic acid.

x 20 salts: 960mM Tris, 160mM Magnesium acetate, 200mM Potassium acetate, 1.26M Ammonium acetate, pH adjusted to 8.0 with glacial acetic acid.

x 20 Energy mix: 50mM ATP, 10mM GTP, 200mM PEP (Na_3salt) pH adjusted to 7.0 with 5M KOH

x 20 amino acid mix: 0.5mM of each of 19 amino acids (minus methionine)

1 x Protein sample buffer: 60mM Tris, 2% (w/v) SDS, 10% (v/v) glycerol, 0.001% (w/v) bromophenol blue, 5% (v/v) 2-mercaptoethanol pH adjusted to 6.8 with HCl

5 x Gel running buffer: 1.875M Tris pH adjusted to 8.9 with 1M HCl

Acrylamide: 30% (w/v) acrylamide, 0.2% (w/v) NN'Methylenebisacrylamide stored in dark at +2°C

Polyacrylamide gel stain: 0.2% (w/v) Coomassie Brilliant Blue R in 40% (v/v) methanol, 7% (v/v) acetic acid

Destain: 40% (v/v) methanol,7% (v/v) acetic acid

Scintillant: 0.4% (w/v) Butyl-PBD in toluene

2.2.3.2 Preparation of E. coli S-30

E. coli S-30 for in vitro translation of heterologous RNA was prepared essentially as described by Modollel (1971).

2.2.3.2.1 Growth of E. coli A single colony of E. coli CSH 73 Hfr H Δ lac Δ (ara-leu) (Miller 1972) was picked off an L agar plate and used to inoculate a 5ml overnight culture. This was diluted into 4 x 100ml cultures which were grown overnight with shaking at 37°C. 50mls of the overnight cultures were transferred into each of 8 x 2 litre flasks containing 500mls L broth. The cultures were incubated with vigorous shaking until mid-log phase (A_{650} 0.6). Growth was terminated by cooling the culture on ice. Cells were harvested by centrifugation at 12,000xg_{ave} for 5 minutes at 4°C and washed 3 times with Buffer A. Unless the cells were to be used immediately they were quick frozen and stored at -80°C.

2.2.3.2.2 Preparation of E. coli S-30 extract E. coli cells (5 - 7g) were broken by grinding with twice the cell weight of

alumina until the mixture was homogeneous and emitted 'popping sounds' (Nirenberg and Matthei, 1961). The paste was extracted by grinding for 2 minutes in 1.5 vols of Buffer A per g cells. Alumina, whole cells and debris were pelleted by centrifugation at 20,000xg_{ave} for 30 minutes. The upper 80% of the supernatant was removed and further clarified by centrifugation at 30,000xg_{ave} for 30 minutes.

The upper 90% of clarified supernatant was transferred to a graduated test tube. For each ml of supernatant the following was added:

100	µl	1M Tris acetate pH 8.2
2.5	µl	1M magnesium acetate
14	µl	50 mM ATP
2	µl	50 mM GTP
40	µl	200 mM PEP
1	µl	Pyruvate Kinase (2 units)
20	µl	0.5mM 19 amino acids (less methionine)
20	µl	0.5mM methionine
0.8	µl	1M DTT

Endogenous mRNA activity was reduced by incubating the extract at 37°C in darkness for 2 hours. The preincubated extract was then dialysed against several changes of Buffer A for 24 hours at 4°C, centrifuged at 30,000xg_{ave} for 30 mins. and 200µl aliquots of the supernatant quick frozen in liquid N_2 and stored at -80°C.

2.2.3.3 Conditions of translation by E. coli S-30 extract

The conditions for translation of RNA by the E. coli S-30 extract were essentially as described by Bottomley and Whitfeld (1979) with a few modifications (Walden 1978).

2.2.3.3.1 Translation of exogenous mRNA In vitro translation reactions were carried out in a final volume of 50µl in 1.5ml polypropylene tubes. RNA (0 - 40µg) was diluted to 30µl with sterile double distilled water. To each incubation was added 20µl of S-30 reaction mix made up as follows:

S-30 reaction mix components per incubation

x 20 amino acids	2.5µl
x 20 salts	2.5µl
x 20 energy	2.5µl
1M DTT	0.1µl
Pyruvate Kinase	0.5 unit
^{35}S methionine	4 - 10µ Ci
H_2O to	20µl

The final concentrations of the reagents in each incubation including the contribution from the E. coli S-30 were as follows: 50mM Tris acetate, 10mM magnesium acetate, 75mM ammonium acetate, 10mM potassium acetate, 10mM PEP, 0.5mM GTP, 2.5mM ATP, 0.5 unit Pyruvate Kinase, 0.025mM each 19 amino acids, 2mM DTT, 0.3-2.0µM ^{35}S methionine. The reactions were incubated at 37°C for the required length of time (up to 40 minutes). An estimate of ^{35}S methionine incorporated into protein was made by spotting duplicate 5µl aliquots of each incubation onto 1cm squares of Whatman 3MM paper which were washed as follows:

1. 10 minutes wash in 100mls 10% (w/v) TCA, 0°C
2. 15 minutes wash in 100mls 5% (w/v) TCA, 95°C
3. 4 x 5 minutes washes in 100mls 5% (w/v) TCA, room temperature
4. 15 minutes in 100mls 50% (v/v) ethanol, 50% (v/v) diethyl ether
5. 15 minutes in 100mls 100% ether, 37°C

Filters were then dried and the radioactivity estimated in 5mls of scintillant in an Intertechnique scintillation counter. An estimate of background radioactivity was made by spotting 5µl of S-30 "reaction mix" onto Whatman 3MM filters, which were washed as described above. Total available cpm were estimated by counting filters onto which 5µl of S-30 reaction mix had been spotted, but which had not been washed in TCA.

Incorporation of ^{35}S methionine into protein by the S-30 was terminated by adding 1ml 80% (v/v) acetone, and protein precipitated by standing the tubes on ice for 30 mins. The precipitated protein was pelleted by centrifugation at 12,000xg_{max} for 3 minutes and then dried under vacuum.

2.2.3.3.2 Coupled transcription/translation of exogenous DNA

The conditions for *in vitro* coupled transcription/translation of exogenous DNA were identical to those described for translation (section 2.2.3.3.1) except

1. CTP and UTP were included at a final concentration of 0.5mM each
2. Each reaction (50µl) generally contained 2.5µg of plasmid or viral DNA. Incubations and estimations of incorporation of ^{35}S methionine into protein were identical to those described in section 2.2.3.3.1.

2.2.3.4 SDS polyacrylamide gel electrophoresis of proteins

Polypeptides synthesised by the *E. coli* S-30 were analysed by electrophoresis on SDS polyacrylamide slab gels (Laemmli and Favre 1973).

Vertical gels were cast between 2 glass plates held apart by perspex spacers to give a gel mould 220mm long x 140mm wide x 1mm thick.

A 15% slab gel contained the following:

15% (w/v) acrylamide

0.1% (w/v) NN'Methylenebisacrylamide

0.375M Tris Cl pH 8.9

0.1% SDS

0.04% (v/v) TEMED

0.05% (w/v) Ammonium persulphate

The above gel components with the exception of TEMED and SDS were mixed and degassed. TEMED and SDS were then added and the gel poured. The polymerised slab gel was overlain with a 5% stacking gel containing the following:

5% (w/v) acrylamide

0.033% (w/v) NN'Methylenebisacrylamide

60mM Tris HCl pH 6.8

0.1% (w/v) SDS

0.1% (w/v) TEMED

0.04% (w/v) Ammonium persulphate

Sample wells (17mm long x 7.5mm wide x 1mm thick) were formed with a 12 slot comb.

Once the stacking gel had polymerised the bottom spacer was removed and the apparatus clamped to a vertical gel stand. The buffer tanks contained 1 x gel running buffer, the upper tank also containing 0.01% SDS. The precipitated proteins from the S-30 incorporation were re-suspended in 1 x protein sample buffer (approx. 1µl buffer per 10,000 cpm ^{35}S incorporated into protein). Samples were boiled for 2 minutes, and approximately 20µl loaded onto the gel. Electrophoresis was carried out overnight at 9 mA. Marker proteins were run in parallel tracks to provide reference molecular weights. These were:

Bovine Serum Albumin (M_r 68,000)

Catalase (M_r 60,000)

Aldolase (M_r 40,000)

Carbonic anhydrase (M_r 29,000)

Soybean trypsin inhibitor (M_r 21,000)

Myoglobin (M_r 17,000)

Marker proteins were visualised by staining in 200mls gel stain for two hours. Background stain was removed by soaking the gel in gel destain for at least six hours. Gels which were to be autoradiographed were dried under vacuum onto Whatman 3MM paper using a slab gel dryer (Biowerk Basel) on high heat setting, covered with 'Saran wrap' (Dow) and exposed to X ray film (180mm x 240mm, Cronex) at room temperature for 2 - 21 days.

2.2.4 Gel electrophoresis of RNA and Northern blotting

RNA electrophoresed through agarose/formaldehyde denaturing gels (Lehrach et al 1977) modified by Maniatis et al (1982) was blotted to nitrocellulose and hybridised with DNA probes by the method of Thomas (1980).

2.2.4.1 Buffers

10 x MOPS: 0.2M MOPS, 50mM sodium acetate, 10mM EDTA pH adjusted to 7.0 with 10M NaOH

2.5 x FSB: 50mM MOPS, 47% (v/v) formamide, 11% (v/v) formaldehyde

5 x FDE: 0.1M EDTA pH 8.0, 30% (w/v) Ficoll type 400,000, 0.01% (w/v) Bromophenol blue

100 x Denhardt's solution (Denhardt (1966): 2% (w/v) Ficoll type 400,000, 2% (w/v) PVP, 2% (w/v) BSA in 3 x SSC

2.2.4.2 Electrophoresis of RNA through agarose/formaldehyde gels

RNA was normally fractionated by electrophoresis through 1.3% (w/v) horizontal agarose gels. 1.95g of agarose was boiled with 109mls water for five minutes until completely dissolved. The agarose was cooled to 60°C then 15mls 10 x MOPS and 26mls 40% (v/v) formaldehyde solution added. The gel was cast into a mould to give final dimensions of 14mm wide x 28mm long x 5mm thick. Slots 1mm thick x 7mm wide and 10mm deep were formed by a 13-tooth comb. The gel bed was supported on 2 tanks each containing 500mls 1 x MOPS, and electrical continuity provided by 3-ply 'J cloth' wicks (Chicopee Ltd., Gwent). Samples (maximum volume 10µl) were mixed with an equal volume of 2.5 x FSB and the RNA denatured by heating at 60°C for 5 minutes in a 1.5ml polypropylene tube. The RNA was snap cooled on ice, mixed with 0.25 vol of 5 x FDE and loaded immediately onto the gel. Generally 5 - 10µg RNA was applied to each track.

Electrophoresis was at 40mA for 15 minutes to load the samples, after which the buffer in the wells was topped up and the whole gel covered with plastic film.

Electrophoresis was continued at 10mA for 15 - 20 hours until the bromophenol blue had migrated to the end of the gel.

2.2.4.3 Transfer of RNA to nitrocellulose

Immediately after electrophoresis the gel was placed on a wick made from three thicknesses of Whatman 3MM paper soaked in 20 x SSC. The wick was suspended on a glass plate between two tanks each containing 200mls 20 x SSC. Care was taken to extrude all air bubbles between gel and wick. A piece of nitrocellulose (0.45µm pore size, Schleicher and Schuell) was cut to fit the gel, wetted with water and

equilibrated with 20 x SSC for 10 minutes before transfer. The nitrocellulose was gently lowered onto the gel, again ensuring no air bubbles were trapped. The top edge of the nitrocellulose was precisely aligned with the gel slots to allow accurate sizing of transcripts. The nitrocellulose was covered with six sheets of Whatman 3MM paper soaked in 20 x SSC, followed by a four-inch stack of 'Wyp-All' (Bowater Scott) absorbent tissue. A 1kg weight was applied to the blot and the apparatus covered in plastic film. Transfer was allowed to proceed for at least 18 hours. The nitrocellulose was then dried under a 250 W infra red lamp and baked in a vacuum oven for 2 hours at 80°C. Prior to use filters were stored at +2°C between 2 sheets of Whatman 3MM paper and sealed in polythene.

2.2.4.4 Staining and photography of gels

RNA was visualised by the method of Maniatis et al (1982). Since staining of RNA prior to transfer to nitrocellulose reduces the efficiency of transfer by up to 50% (Thomas 1980), RNA size markers were run in parallel tracks to the RNA which was to be blotted, and these excised and stained separately. Formaldehyde was removed by gently shaking the gel for one hour in several volumes of distilled water with two changes. RNA was renatured by shaking for one hour in two changes of 0.1M ammonium acetate, and stained by further agitation in 0.1M ammonium acetate, 5µg/ml ethidium bromide for 30 mins. Background fluorescence was removed by destaining the gel for 40 minutes in several volumes of 0.1M ammonium acetate.

Gels were illuminated with short wave U.V. light (254nm, Ultraviolet Products Inc.) and photographed through a Kodak Wratten gelatin

filter No.234 on Kodak 2415 technical pan film.

2.2.4.5 Hybridisation of labelled DNA probes to RNA immobilised on nitrocellulose

All prehybridisations and hybridisations were carried out in sealed polythene bags or borosilicate glass scintillation vials in an incubator fitted with a rocking table (Laboratory Thermal Equipment).

2.2.4.5.1 Prehybridisation Filters were prehybridised (Denhardt 1966) for at least one hour and up to 18 hours at 42°C in 50% (v/v) formamide, 5 x Denhardt's solution, 0.1% (w/v) SDS, 250µg/ml sheared and denatured herring sperm DNA, 5 x SSC with 2 - 3 mls buffer/cm^2 paper.

2.2.4.5.2 Hybridisation Hybridisation was carried out in the same buffer used for prehybridisation to which was added approximately 1µg of thermally denatured probe. (Heated for 10 minutes at 100°C and snap cooled on ice). Filters were hybridised for a minimum of 18 hours. Optimal signal to noise ratios were obtained when the probe used was ^{32}P labelled to a specific activity of greater than 2.2×10^6 dpm/µg DNA (section 2.2.10).

2.2.4.5.3 Washing of filters Filters were washed for 2 x 10 minutes with 250mls of 2 x SSC, 0.1% (w/v) SDS, followed by 2 x 60 minutes with 250mls 0.1 x SSC, 0.1% (w/v) SDS irrespective of filter size.

2.2.4.6 Autoradiography and fluorography

^{32}P labelled DNA probes hybridising to RNA immobilised on nitrocellulose were visualised by autoradiography or fluorography using X-ray film.

The nitrocellulose blots were mounted on card, covered with 'Saran wrap' and exposed to DuPont Cronex X-ray film in a metal casette. Exposure was at room temperature for 16 hours. Weaker signals were detected by fluorography; the X-ray film was pre-flashed (Laskey and Mills 1975) and exposed to X-ray film in a metal casette fitted with 'LightningPlus' intensifying screens (Dupont) at -80°C (Randerath 1970) for up to fourteen days.

2.2.5 S1 Nuclease mapping of mitochondrial RNAs

The technique described by Berk and Sharp (1977) was used with some minor modifications (Weaver and Weissmann 1979), for accurately mapping transcripts to DNA.

2.2.5.1 Buffers

Formamide: recrystallised three times at 0°C (Robberson et al 1977)

Hybridisation buffer (Casey and Davidson 1977): 80% (v/v) formamide, 0.4M NaCl, 40mM PIPES (Na_2 salt), 1mM EDTA pH adjusted to 6.4 with 1M NaOH. All components except formamide mixed and autoclaved. Formamide added just before use.

S1 buffer (Vogt 1973): 0.28M NaCl, 0.05M sodium acetate, 4.5mM $ZnSO_4$ pH adjusted to 4.6 with glacial acetic acid

S1 stop buffer: 4.0M ammonium acetate, 0.1M EDTA

2.2.5.2 Preparation of labelled DNA

Cloned fragments of DNA were labelled by one of two methods. Purified restriction fragments were labelled at their 5' ends using T4 polynucleotide kinase (section 2.2.10.3). DNA fragments cloned in M13 (where the sequence of the strand to be labelled was complementary to the RNA) were labelled by second strand synthesis (section 2.2.10.4) except that 16µl of template DNA (estimated to be

1µg) was annealed with 2µl primer, and 30 - 40 µCi α ^{32}PdCTP was included in each incubation

2.2.5.3 Hybridisation conditions

DNA (10 - 500ng, at least 10,000 dpm) was co-precipitated together with 15 - 25µg mt RNA at -80°C for 30 minutes in a 1.5ml polypropylene tube, by adding $^{1}/_{10}$ volume 3M sodium acetate pH 5.5 and 3 volumes of ethanol. DNA and RNA were pelleted at 12,000xg_{max} for 10 minutes at 4°C, washed twice with 70% (v/v) ethanol, desiccated and resuspended in 30µl hybridisation buffer. DNA and RNA were dissolved in the hybridisation buffer by repeated pipetting with a micropipette. Solubilisation of nucleic acid (which took up to one hour) was checked by monitoring Cerenkov counts in solution. The tubes were then sealed with Parafilm (American Can Co.) and submerged in a water bath at 82°C for 15 minutes to ensure strand separation of duplex DNA and denaturation of RNA secondary structure. Hybridisation was initiated by quickly transferring the tubes to water baths at the required temperature. The temperature of hybridisation was selected so as to be 5 - 10°C higher than the melting temperature (T_m) of the mt DNA. In high concentrations of formamide, Casey and Davidson (1977) have shown that the T_m of DNA:DNA duplex is 5 - 10°C lower than the T_m of the corresponding DNA:RNA. The T_m of the DNA was calculated from the formula

$$T_m \ ^{\circ}C = 81.5 + 0.41(GC) - 0.72F + 16.6\log_{10}M$$

(Marmur and Doty 1962, McConaughy et al 1969, Howley et al 1979, Casey and Davidson 1977) where

(GC) is the % GC content of the restriction fragment or clone used

F is the % formamide in the hybridisation buffer

M is the monovalent cation concentration in the hybridisation buffer. Tubes were incubated submerged in the water bath(s) for 3 hours. Hybridisations were terminated by adding 300µl ice cold S1 buffer containing 100 - 400 U/ml S1 nuclease to each tube, keeping as much of the tube submerged as possible. (One unit of S1 nuclease hydrolyses 1µg of denatured DNA in 1 minute at 37°C). The samples were incubated with S1 nuclease at 37°C for 20 minutes, after which time 50µl of S1 stop buffer was added, and the whole phenol extracted. DNA:RNA S1 protected hybrids were precipitated with 20µg yeast tRNA at -20°C overnight by the addition of 2.5 volumes of ethanol. Precipitated DNA:RNA hybrids were pelleted by centrifugation at 12,000xg_{max} for 30 minutes, washed once with 70% (v/v) ethanol and desiccated. The samples were then resuspended in 3µl TE80 and 2µl of Sanger dye (section 2.2.12.1) was added to each tube.

2.2.5.4 Electrophoresis of DNA:RNA hybrids

S1 nuclease protected fragments were electrophoresed on 8M Urea/6% (w/v) polyacrylamide gels exactly as described in section 2.2.12.4. Size markers were provided either by labelled restriction fragments [ØX174 plus Bsp R1 1342, 1078, 871, 606, 310, 276, 271, 234, 194, 118 and 72bp (M. de Haan pers. commun)] or by a well characterised M13 clone sequenced by the Sanger chain termination method (section 2.2.12.3). In the latter case it was necessary to know the distance in bp from the 3' end of the primer to the cleavage point of the restriction site used for cloning, in addition to the length of insert sequence. Gels were run for 2½ hours (for S1 protected fragments up to 100bp) or up to 6 hours for fragments up to 300bp. Gels were fixed, dried and autoradiographed for 1 - 3 days as described in section 2.2.12.4.

2.2.6 Preparation of plasmid DNA

Plasmid DNA was purified by the detergent lysis method [Godson and Vapneck (1973) modified by Hu and Messing (1982)]

2.2.6.1 Buffers

Ampicillin 100mg/ml in water

Sucrose mix: 25% (w/v) sucrose, 50mM Tris, 40mM EDTA pH adjusted to 8.1 with 1M HCl

Triton mix: 0.1% (w/v) Triton X-100, 50mM Tris, 62.5mM EDTA pH adjusted to 8.1 with 1M HCl

2.2.6.2 Bacterial host strain

E. coli strain HB101 (F^-, hsd S20, (r^-_B, m^-_B), recA13, ara-14, proA2, lacY1, galK2, rpsL20(Sm^r), xyl-5 mtl-1, supE44, λ^-) (Bolivar and Backman 1979) was used for the propagation of plasmid clones.

2.2.6.3 Plasmid vectors

Three plasmid vectors containing mt DNA clones were used. pBR322 (Bolivar et al 1977) was used as a vector for the cytochrome c oxidase subunit II gene (Fox and Leaver 1981). Apocytochrome b clones were maintained in pBR328 (Soberon et al 1980), and pAT153 (Twigg and Sherratt 1980) was the vector for cytochrome c oxidase subunit I clones. All recombinant clones were Ampicillin resistant.

2.2.6.4 Method

A 'stab' culture of HB101 containing the desired plasmid was streaked on a plate of L agar containing Ampicillin at 100µg/ml to generate single colonies. A single bacterial colony was used to inoculate a 5ml 'overnight' culture containing 2 µg/ml Ampicillin. This culture

was diluted into 500mls of L broth, (containing 0.2% (w/v) glucose, 20µg/ml Ampicillin) and grown in a 2½ litre flask at 37°C with shaking in a New Brunswick orbital shaker (250rpm) until A_{590} 1.0. Solid chloramphenicol was added to 150µg/ml and the cultures incubated for a further 18 hours to allow plasmid amplification (Clewell 1972). Cells were harvested by centrifugation at 4,000xg_{ave} for 15 minutes in a Sorvall GSA rotor and resuspended in 6mls sucrose mix. To this was added 1ml 10mg/ml lysozyme freshly dissolved in sucrose mix, and 1ml 0.5M EDTA pH 8.1. Digestion was carried out for 5 minutes on ice. The cells were lysed by the addition of 13mls Triton mix and allowed to stand for a further 15 minutes on ice. At this stage the mixture became very viscous. The lysate was clarified by centrifugation at 48,000xg_{ave} for 45 minutes at 4°C. The supernatant was decanted into a measuring cylinder and 0.95g CsCl and 0.1ml 5mg/ml ethidium bromide added per ml of supernatant. The CsCl solution was transferred to 2 x 13.5ml polyallomer tubes and centrifuged to equilibrium for 48 - 65 hours at 100,000xg_{ave} at 18°C (Radloff et al 1967). After centrifugation the tubes were viewed under U.V. light (366nm, Ultraviolet Products Inc.) and the lower fluorescent band containing supercoiled plasmid DNA withdrawn with a 19G needle attached to a sterile syringe. Ethidium bromide was removed by repeated extraction with an equal volume of 1:1 butan-1-ol:iso propanol equilibrated with CsCl-saturated TE80 until no more colour remained in the aqueous phase. The DNA was then dialysed for 24 hours at 4°C against several changes of TE80, ethanol precipitated, washed twice with 70% (v/v) ethanol, and resuspended in 100µl of H_2O. Estimates of the yield of plasmid DNA were made by measuring A_{260} as described in section 2.2.1.4.

2.2.7 Digestion of DNA with restriction endonucleases

2.2.7.1 Buffers

10 x low salt: 10mM Tris, 10mM $MgCl_2$, 1mM DTT pH adjusted to 7.5 with 1M HCl

10 x medium salt: 50mM NaCl, 10mM Tris, 10mM $MgCl_2$, 1mM DTT pH adjusted to 7.5 with 1M HCl

10 x high salt: 100mM NaCl, 50mM Tris, 10mM $MgCl_2$, 1mM DTT pH adjusted to 7.5 with 1M HCl

Medium to High salt conversion buffer: 0.5 M NaCl, 0.4M Tris pH adjusted to 7.4 with 1M HCl. 1µl added to 10 µl reaction volume

2.2.7.2 Reaction conditions

1 - 5µg DNA was generally digested with the appropriate restriction endonuclease in a final volume of 20µl. Larger amounts of DNA were digested in proportionately increased volumes. Reactions contained 0.1 volume of the appropriate 10 x buffer (section 2.2.7.1 and Table 2.1) and 5 - 20 U of restriction enzyme. Plasmid DNA and gel-purified restriction fragments were usually digested to completion with a 5-fold excess of enzyme over the manufacturers' recommendations. Reactions were incubated at 37°C for 1 - 2 hours except TaqI which was incubated at 65°C. TaqI digestions were overlain with a thin layer of paraffin to prevent evaporation. Where two restriction endonucleases with different buffer requirements were used to digest DNA, the enzyme with the lower salt requirement was used first. After 1 hour the second enzyme was added with 2µl of conversion buffer per 20µl reaction volume. Enzymes were inactivated by heating at 68°C for 10 minutes followed by snap cooling on ice, except TaqI which was phenol extracted and the DNA precipitated with ethanol.

Table 2.1 Buffer requirements of restriction endonucleases used

Low Salt	Medium Salt	High Salt
MspI	AccI	EcoRI
TaqI	BamHI	
	HindIII	
	HinfI	
	Sau3A	

2.2.8 Gel electrophoresis of DNA

2.2.8.1 Buffers

20 x TAE: 0.8M Tris, 0.4M sodium acetate, 0.02M EDTA, pH adjusted to 8.2 with glacial acetic acid

Agarose gel loading buffer: 80mM EDTA, 60% (w/v) sucrose, 0.3% (w/v) Bromophenol blue stored at +2°C

Polyacrylamide gel loading buffer: 25% (v/v) glycerol, 0.3% (w/v) Bromophenol blue, 0.3% (w/v) xylene cyanol FF in 10 x TBE stored at room temperature.

2.2.8.2 Agarose gel electrophoresis of DNA

1% (w/v) agarose gels were used to separate DNA fragments of $>$ 1Kb and up to 10Kb. 0.8% (w/v) gels were run where DNA fragments $>$ 2.5Kb and of similar size needed to be separated. The appropriate amount of agarose (Miles for routine separations, Sea Kem for preparative isolations) was mixed with 150mls 1 x TAE to give the desired final agarose concentration. The agarose was boiled for ten minutes and then cast in the same mould as described for electrophoresis of RNA (section 2.2.4.2). DNA sample volumes were adjusted to 20µl, mixed with 5µl of loading buffer and applied to the gel. Electro-

phoresis was at 7.5V/cm until the Bromophenol blue had migrated into the gel. The voltage was then reduced to 1.4V/cm and electrophoresis continued overnight, until the dye had migrated to within 5cm of the end of the gel. Plastic film was placed over the gel to prevent dehydration during electrophoresis.

2.2.8.3 Submerged agarose mini-gels

Submerged agarose mini-gels provided a rapid means of analysing small quantities of DNA, most frequently to assess the progress of a restriction enzyme digest and also to quantify yields of DNA. 1% (w/v) agarose gels in 1 x TBE were cast in moulds to give final dimensions of 100mm long x 69mm wide x 3mm deep. A miniature comb was used to form sample wells with dimensions 5mm wide x 5mm deep x 1mm thick. The gel was submerged in 30mls 1 x TBE. Sample volumes were adjusted to 8µl with water, mixed with 2µl loading buffer and applied to the gel (up to 1µg per track). Electrophoresis was at 50mA for 45 minutes or until the dye had migrated to the end of the gel.

2.2.8.4 Polyacrylamide gel electrophoresis of DNA

Polyacrylamide gels were used to separate and prepare DNA fragments of $<$ 1Kb (Maniatis et al 1975).

Vertical gels were cast between two glass plates held apart by perspex spacers as described in section 2.2.3.4. A 6% polyacrylamide gel contained the following:

5.8% (w/v) acrylamide	0.05% (w/v) TEMED
0.2% (w/v) NN'Methylenebisacrylamide	5% (v/v) glycerol
0.07% (w/v) ammonium persulphate	1 x TBE

A stock solution of 30% (w/v) acrylamide (29:1 acrylamide: NN'Methylene bisacrylamide) was stored in the dark at +2°C. Ammonium persulphate was freshly prepared. The above gel components with the exception of TEMED were mixed and degassed. TEMED was then added and the gel poured. Sample wells were formed with a 12 slot comb (final sample well dimensions 7.5mm wide x 17mm deep x 1mm thick). Immediately the gel had polymerised the comb was removed and the wells rinsed thoroughly with 1 x TBE. The bottom spacer was removed and the gel clamped to a vertical electrophoresis stand with 500mls 1 x TBE in each tank. Gels were pre-run for 60 minutes before loading to remove TEMED and AMPS from the gel.

DNA samples (up to 5µg/track, generally 20µl) were mixed with 0.25 volume of loading buffer and applied to the gel using a Hamilton syringe. Electrophoresis was at 10V/cm for 2½ hours.

2.2.8.5 Staining of gels

DNA was visualised in agarose and polyacrylamide gels by staining with ethidium bromide (Sharp et al 1973). Gels were immersed in water containing 0.5µg/ml ethidium bromide for 30 minutes and de-stained in water for 15 minutes, then photographed as described in section 2.2.4.4.

2.2.9 Recovery of DNA from gels

2.2.9.1 Buffers

Maxam and Gilbert gel elution buffer (for polyacrylamide gels):

0.5M ammonium acetate, 10mM Tris, 1mM EDTA pH adjusted to 7.5 with 1M HCl

Electroelution buffer (for agarose gels): 5mM Tris base, 2.5mM acetic acid

2.2.9.2 Recovery of DNA fragments from polyacrylamide gels

DNA restriction fragments were recovered from polyacrylamide gels by the method of Maxam and Gilbert (1977). DNA fragments were separated in 6% (w/v) polyacrylamide gels as described in section 2.2.8.4, stained with ethidium bromide and viewed on a U.V. trans-illuminator (302nm, Ultra Violet Products Inc.). The desired DNA band was excised from the gel using a sterile scalpel and crushed to a fine pulp with 0.5ml gel elution buffer in a 1.5ml polypropylene tube using a Teflon plunger. An additional 0.5ml elution buffer was added and the DNA eluted overnight at room temperature. DNA fragments which were to be 5' labelled with kinase (section 2.2.10.3) were eluted in the same volume of TE80 at 4°C overnight. The eluted DNA was separated from the acrylamide by centrifuging at 1,000xg_{ave} for 5 minutes through a glass wool plug in a 5ml syringe. The glass wool pad was washed with 0.5ml elution buffer/TE80. The DNA was precipitated at -80°C for 30 minutes with 3 volumes of ethanol and pelleted by centrifugation at 18,000xg_{ave} for 25 minutes. The resulting DNA pellet was resuspended in a small volume of TE80.

2.2.9.3 Electroelution of DNA fragments from agarose gels

DNA fragments were recovered from agarose gels by electroelution (McDonell et al 1977). DNA fragments were separated by electrophoresis through agarose gels as described in section 2.2.8.2, stained with ethidium bromide and viewed on a transilluminator (302nm, Ultra Violet Products Inc.). The DNA band to be eluted was excised from the gel in as small a volume of agarose as possible using a sterile scalpel. The gel slice was placed in a piece of sterile pre-treated (Maniatis et al 1982) dialysis tubing together with

approximately 2mls of electroelution buffer. The tubing was sealed at both ends and immersed in a tank containing electroelution buffer. A potential difference was passed across the tank (10V/cm) with the gel slice perpendicular to the current. The progress of electroelution was followed by viewing the gel slice on a U.V. transilluminator until no more ethidium bromidefluorescence was seen in the gel slice. When electroelution was complete the polarity of the current was reversed for 10 minutes. The buffer containing eluted DNA was removed from the dialysis tubing and extracted with an equal volume of phenol. The phenol was back extracted with 1ml electroelution buffer.

Ethidium bromide was removed by washing with an equal volume of n-butanol. DNA was precipitated at -80°C for 30 minutes by the addition of sodium acetate to 0.3M and 3 volumes of ethanol. The precipitated DNA was pelleted at 18,000xg_{ave} for 30 minutes, washed twice with 70% (v/v) ethanol and resuspended in a small volume of TE80.

The DNA prepared by this method was found to contain impurities which inhibited restriction enzymes. Thus DNA electroeluted from agarose gels was routinely dialysed for 24 hours at +4°C against several changes of TE80, and reprecipitated.

2.2.10 ^{32}P labelling of DNA

2.2.10.1 Buffers

1 x Nick translation buffer: 52.5mM Tris, 5.25mM $MgCl_2$, 20μM each d(A G C and T)TP, 35mM 2-mercaptoethanol pH adjusted to 7.5 with 1M HCl

DNase dilution buffer: 22.5mM Tris, 45mM $(NH_4)_2SO_4$, 0.045% (w/v) BSA, 4.5mM 2-mercaptoethanol, 50% (v/v)glycerol pH adjusted to 7.5 with 1M HCl

'Klenow' buffer: 5.25mM Tris, 0.1mM EDTA, 75µM each d(A G T)TP pH adjusted to 8.0 with 1M HCl

10 x CIP buffer: 0.5M Tris, 10mM $MgCl_2$, 1mM $ZnCl_2$, 10mM spermidine pH adjusted to 9.0 with 1M HCl

10 x Kinase buffer: 0.5M Tris, 0.1M $MgCl_2$, 50mM DTT, 1mM EDTA, 1mM spermidine pH adjusted to 7.6 with 1M HCl

2.2.10.2 Nick Translation of DNA

Plasmid DNA and DNA restriction fragments were ^{32}P labelled by nick translation (Rigby et al 1977). Nick translation reactions were carried out in a final volume of 33µl. The following reagents were mixed:

DNA in TE80 1µg

α ^{32}P d CTP 10 - 30 µCi

2 x 10^{-5}mg/ml DNase I

(freshly diluted in DNase I diluting buffer) 2µl

DNA pol I 1 unit

1 x nick translation buffer to 33µl

The reaction was incubated for 90 minutes at 15°C and terminated by the addition of 50µl TE80. Labelled DNA was separated from unincorporated nucleotides by centrifugation through a small column of Sephadex G50 (section 2.2.10.5).

2.2.10.3 Labelling 5' ends of DNA restriction fragments

Protruding termini of DNA restriction fragments which had been dephosphorylated were ^{32}P labelled using T4 polynucleotide kinase

and $\gamma\ ^{32}P$ d ATP.

2.2.10.3.1 Dephosphorylation of DNA Terminal 5' phosphates were removed from gel purified DNA restriction fragments using CIP (Chaconas and van de Sande 1980). The following reagents were mixed:

DNA	500ng
10 x CIP buffer	2µl
CIP	2 units
H_2O	to 20µl

The reaction was incubated at 37°C for 1 hour and then at 68°C for 30 minutes to inactivate the enzyme. CIP was removed by phenol extraction. The phenol was back extracted with TE80, the aqueous phases pooled and DNA precipitated overnight at -20°C by adding 0.1 volume 3M sodium acetate and 3 volumes of ethanol. Precipitated DNA was pelleted by centrifugation at 12,000xg_{ave} for 10 minutes, washed once with 70% (v/v) ethanol, desiccated and dissolved in 5µl TE80.

2.2.10.3.2 Labelling 5' ends with T4 polynucleotide kinase

The dephosphorylated restriction fragments were labelled at their 5' termini using $\gamma\ ^{32}P$ d ATP and T4 polynucleotide kinase (Maxam and Gilbert 1980). The following reagents were mixed:

dephosphorylated DNA	500ng
10x Kinase buffer	2µl
$\gamma\ ^{32}P$ d ATP	50 µCi
T4 polynucleotide kinase	10 units
H_2O	to 20µl

The reaction was incubated at 37°C for 30 minutes and then stopped by the addition of the following:

75µl 100mM Tris Cl, 1mM EDTA pH 7.0

2.5µl 10% (w/v) SDS

2.5µl 500mM EDTA pH 7.5

The labelled DNA was separated from unincorporated γ^{32}P d ATP as described in section 2.2.10.5. The efficiency of labelling of 5' ends of DNA restriction fragments was generally ca. 0.5%.

2.2.10.4 Second strand synthesis of M13 bacteriophage DNA

Single stranded DNA clones in bacteriophage M13 were labelled to high specific activity by a modification of the sequencing reaction (section 2.2.12 and Sanger et al 1977) using DNA polymerase I (Klenow fragment). 8µl of template DNA was annealed with 1µl of sequencing primer as described in section 2.2.12.2. 9µl of Klenow buffer and 1µl of α^{32}P d CTP (10µCi) were then added, followed by 2.5 units Klenow freshly diluted to 10µl in 10mM Tris HCl pH 8.0. After incubating for 1 hour at 30°C the reaction was terminated by the addition of 50µl TE80. Incorporated nucleotides were separated from free label as described in section 2.2.10.5.

2.2.10.5 Separation of ^{32}P-labelled DNA from unincorporated nucleotides

^{32}P-labelled fragments of DNA were separated from unincorporated nucleotides by centrifugation through a small column of Sephadex G50. A 1ml plastic hypodermic syringe was plugged with glass wool and filled with sterile Sephadex G50 (fine) pre-swollen and equilibrated with TE80. The column was dried by centrifugation at 500xg_{ave} for 1½ minutes which reduced the packed volume to 0.7ml. The DNA mixture was applied to the top of the column and recentrifuged at 750xg_{ave} for 2 minutes. The eluate containing labelled DNA was collected in a 1.5ml polypropylene tube; unincorporated label remained

on the column.

2.2.10.6 Determination of the efficiency of DNA labelling

The efficiency of ^{32}P dNTP incorporation into DNA was estimated by measuring Cerenkov light emitted by ^{32}P in aqueous solution (Marshall 1952). Counts per minute were converted to disintegrations per minute using a calibration curve constructed for the Inter-technique SL-3000 scintillation counter. This curve was constructed using ^{32}P of known activity (data not shown) since 1μCi ^{32}P = 2.2×10^6 dpm. Decay of the isotope was accounted for by the correction factor $2^{-t/t_{\frac{1}{2}}}$ where t = days after activity date, $t_{\frac{1}{2}}$ = 14.31 days (half life of ^{32}P)

2.2.11 Cloning DNA restriction fragments into bacteriophage M13

2.2.11.1 Introduction

DNA fragments were cloned into replicative form (RF) bacteriophage M13 vector mp8 (Messing and Vieira 1982). These clones were used to prepare single stranded template DNA (Messing et al 1977) from which high specific activity ^{32}P labelled single stranded DNA probes were generated (section 2.2.10.4). Single stranded M13 probes were hybridised with RNA for Northern blotting and S1 nuclease mapping when the DNA insert was oriented such that the strand labelled with Klenow was complementary to the RNA.

2.2.11.2 Buffers and indicator solutions

LTB: 20mM Tris, 20mM NaCl, 1mM EDTA pH adjusted to 7.9 with 1M HCl

PEG-NaCl: 2.5M NaCl, 20% (w/v) polyethyleneglycol 6000

TE(0.1)80: 10mM Tris, 0.1mM EDTA pH adjusted to 8.0 with 1M HCl

IPTG: 14mg/ml IPTG in dimethyl formamide, stored at -20°C

X-gal: 20mg/ml X-gal in dimethyl formamide, stored at -20°C.

10 x ligase: 0.66M Tris, 0.1M $MgCl_2$, 0.1M DTT, 1mM ATP, 10mM EDTA pH adjusted to 7.2 with 1M HCl. Stored at -20°C.

Lysis buffer: 25mM Tris, 10mM EDTA, 50mM glucose, 0.2% (w/v) lysozyme pH adjusted to 8.0 with 1M HCl. Freshly prepared.

Alkaline-SDS: 0.2M NaOH, 1% (w/v) SDS

High Salt solution: 3M sodium acetate, 30µg/ml yeast tRNA, pH adjusted to 4.8 with glacial acetic acid.

2.2.11.3 Restriction and ligation

DNA restriction fragments were ligated (Tanaka and Weisblum 1975) into compatable restriction sites in the mp8 polyclonal linker, for example MspI and TaqI fragments were cloned into the AccI site.

Approximately 150ng mp8 RF DNA was prepared for ligation by restriction to completion with the appropriate enzyme. The extent of digestion was measured by electrophoresis of the sample on a mini-gel (section 2.2.8.3); digestion was assessed to be complete when closed circular and supercoiled forms had disappeared. The enzyme was inactivated by heat treatment. The DNA to be cloned was prepared in a similar manner.

150ng M13 RF DNA was ligated with sufficient foreign DNA such that ligation mixes contained a 3:1 molar excess of foreign DNA to vector. Ligation mixes contained the following in a final volume of 10µl:

2 - 3µl DNA (vector and foreign DNA)

1µl 10 x ligase

2 units T4 DNA ligase

Ligations were allowed to proceed for 16 hours at +4°C.

2.2.11.4 Bacterial host and preparation of competent cells

E. coli JM101 [(Δ *lac pro*), *thi*, *supE*, F'*traD36*, *proAB*, *lac* i^q, z Δ *M15*, Messing 1979] was used as a host for M13 clones. Competent JM101 were prepared by the standard $CaCl_2$ procedure (Mandel and Higa 1970, modified by Dagert and Ehrlich 1979). A stab culture was streaked onto a minimal agar plate and a single colony used to inoculate a 5ml culture which was grown overnight until stationary. One ml of this was used to inoculate 100ml of L broth which was grown until A_{600} 0.2 - 0.4. The culture was decanted into ice cold 24ml McCartney bottles, chilled on ice for 15 minutes and cells harvested by centrifugation at 1,300xg_{ave} for 7½ minutes. The cells were resuspended in ½ volume of ice cold 50mM $CaCl_2$, and after 20 minutes at 0°C pelleted by centrifugation as above. The pellets were resuspended in $^1/_{10}$ volume ice cold 50mM $CaCl_2$. Competent cells were stored at 4°C and transformed (Cohen *et al* 1973) within four days of preparation. 10µl of 10^{-1} or 10^{-2} dilutions of ligation mix (section 2.2.11.3) were mixed with 200µl of competent cells in a 6ml metal capped glass tube. After 40 minutes on ice, transformation was effected by a heat pulse at 42°C for 2 minutes. The transformed *E. coli* were mixed with 200µl plating cells (*E. coli* JM101 stationary culture diluted 1:100 in L broth and grown at 37°C with shaking for two hours), 30µl X-gal and 20µl IPTG. The mixture was plated onto minimal medium plates with 3.5ml BBL top agar at 45°C, and incubated overnight at 37°C.

2.2.11.5 Plaque purification

Recombinant (white) plaques well separated from other plaques were used to initiate a 1ml mini-prep (section 2.2.11.6). Where plaques were not well separated, plaques were purified further. The desired

plaque was picked off into 50µl of LTB and diluted 10^6 fold with LTB. 200µl of the diluted 'phage were plated onto minimal plates with plating cells etc.as described in section 2.2.11.4. Purification of plaques in this way gave plaques well separated from others, and these were used to initiate mini-preps.

2.2.11.6 Single stranded (ss) M13 DNA purification

Single stranded M13 DNA was prepared by the method of Yamamoto et al (1970) and Schrieir and Cortese (1979).

One ml of an overnight culture of JM101 was diluted into 25ml L broth and grown with shaking at 37°C until A_{650} 0.2. One ml aliquots of the culture were dispensed into 5ml 'bijoux' bottles and infected with M13 'phage picked off from a well separated plaque using a sterile tooth pick. The cultures were incubated for 4½ hours at 37°C with vigorous shaking (Hermann et al 1980) at 300 rpm in a New Brunswick orbital shaker. The cultures were transferred to 1.5ml polypropylene tubes and bacteria pelleted by centrifugation at 12,000xg_{max} for 5 minutes. The bacterial pellet was used to prepare RF DNA (section 2.2.11.7). The supernatants were transferred to clean tubes and 200µl PEG-NaCl added to each. After 30 minutes at room temperature, the 'phage particles were pelleted at 12,000xg_{max} for 5 minutes. The supernatant was carefully removed (traces of PEG 6000 were found to interfere with sequencing and second strand synthesis reactions) and the 'phage particles resuspended in 100µl TE(0.1) 80. DNA was extracted with 50µl of phenol, the aqueous phase adjusted to 0.3M sodium acetate pH 5.5 and precipitated with 2.5 volumes of ethanol at -80°C for 15 minutes. DNA was pelleted at 12,000xg_{max} for 10 minutes, washed with 1ml of ethanol at -20°C and recentrifuged for 5 minutes. The pellet was desiccated,

dissolved in 50µl T(0.1)E80 and reprecipitated as described above. The final pellet was resuspended in 50µl T(0.1)E80 and stored at -20°C.

2.2.11.7 Preparation of RF DNA

RF M13 DNA was purified from the bacterial pellet (section 2.2.11.6) by the method of Birnboim and Doly (1979). The bacterial pellet was resuspended in 100µl of lysis solution and incubated on ice for 30 minutes. The bacterial cells were lysed by incubation on ice for 5 minutes with 200µl alkaline SDS. Chromosomal DNA was selectively precipitated by adding 150µl high salt solution and standing for one hour at 0°C. The lysate was cleared by centrifugation at 12,000xg_{max} for 5 minutes. 400µl of supernatant was removed, transferred to a fresh tube and DNA precipitated by the addition of 1ml ethanol and standing at -20°C for 30 minutes. The DNA was pelleted by centrifugation at 12,000xg_{max} for 10 minutes, resuspended in 100µl 0.1M Na acetate pH 6.0 and reprecipitated. The final pellet was desiccated and dissolved in 10µl TE80.

2.2.12 Sequencing of DNA with chain terminators

Single stranded M13 clones were sequenced by the chain terminating method (Sanger et al 1977 and Sanger et al 1980).

2.2.12.1 Buffers

Primer: 2.5µg/ml synthetic pentadecamer primer (New England Biolabs). Stored at -20°C in 50µl aliquots

10 x anneal buffer: 100mM Tris, 100mM $MgCl_2$ pH adjusted to 8.0 with 1M HCl

Chase buffer: 0.5mM d(AGC and T)TP. Stored at -20°C

Sanger dye: 10mM EDTA, 0.3% (w/v) Bromophenol blue, 0.3% (w/v) xylene cyanol FF, 98% (v/v) formamide. Stored at +4°C

Termination mixes: Between batches there was some variation in the concentrations of dd NTP's required to give optimum results. The concentrations generally used were:

Termination A: 5.2mM Tris, 0.104mM EDTA, 5.43μM dATP, 109μM dGTP 109μM dTTP, 75μM ddATP, pH adjusted to 8.0 with 1M HCl

G: 5.2mM Tris, 0.104mM EDTA, 109μM dATP, 5.43μM dGTP 109μM dTTP, 250μM ddGTP, pH adjusted to 8.0 with 1M HCl

C: 5.2mM Tris, 0.104mM EDTA, 74μM dATP, 74μM dGTP 74μM dTTP, 15μM ddCTP, pH adjusted to 8.0 with 1M HCl

T: 5.2mM Tris, 0.104mM EDTA, 109uM dATP, 109μM dGTP, 5.43μM dTTP, 250μM ddTTP, pH adjusted to 8.0 with 1.0M HCl

Termination mixes were stored at -80°C in 10μl aliquots. The dCTP concentration in each mix was brought to 1.16μM by the addition of the α^{32}P dCTP in the sequencing reaction.

2.2.12.2 Annealing

8μl of ss template DNA (section 2.2.11.6) was mixed with 1μl 10 x anneal buffer and 1μl sequencing primer in a 1.5ml polypropylene tube. The tube was sealed, heated in an oven at 60°C for 45 minutes and cooled to room temperature over 15 minutes. The condensate was collected by brief centrifugation at 12,000xg_{max}.

2.2.12.3 The sequencing reaction

2μl of annealed template was placed on the side of 4 x 1.5ml polypropylene tubes. 2μl of DNA polymerase I Klenow fragment (freshly diluted to 0.25 Unit/μl in 10mM Tris Cl pH 8.0) was added to the side of each tube. One 10μl aliquot of each termination mix per 4 clones

was thawed and mixed with 5μCi α ^{32}P dCTP. 2μl of isotope/termination mix was then added to the side of the tube and the reagents mixed by brief centrifugation at 12,000xg_{max}. After 15 minutes at room temperature, 2μl of chase buffer was added to the side of the tubes which were briefly centrifuged to mix. The chase was allowed to proceed for 15 minutes after which 4μl of Sanger dye was added to each tube. The tube tops were pierced and the samples boiled for 5 minutes to denature the DNA, followed by snap cooling on ice.

2.2.12.4 Sequencing gels

Products of the sequencing reaction were separated on 6% (w/v) polyacrylamide/8M urea gels. Gels were cast between two glass plates held apart by 'plasticard' spacers to give a gel mould 400mm long x 300mm wide x 0.3mm thick. Gels contained the following:

5.7% (w/v) acrylamide

0.3% (w/v) NN'Methylenebisacrylamide

8M Urea

0.06% (w/v) ammonium persulphate

0.08% (v/v) TEMED

1 x TBE

A stock solution of 30% (w/v) acrylamide was stored at -20°C.

10% (w/v) AMPS was freshly prepared.

The gel components with the exception of TEMED were mixed and warmed to dissolve the urea. TEMED was then added and the gel poured. Sample wells were formed with a 45 slot comb made of 0.3mm plasticard. Gels were allowed to polymerise for at least one hour before electrophoresis. The apparatus was clamped to a vertical electrophoresis stand with 500mls 1 x TBE in each tank. Immediately before loading

the sample wells were washed with 1 x TBE to remove urea and unpolymerised acrylamide. Normally approximately ¼ of each sample was loaded. Electrophoresis was generally at 50 - 65 W (power limiting), which maintained the gel temperature sufficiently high to prevent renaturation of DNA strands. Electrophoresis was continued until the Bromophenol blue was just at the bottom of the gel (usually 2 - 2½ hours) - the dye migrates just ahead of the HindIII site in mp8.

Following electrophoresis the gel was fixed by immersion in 10% (v/v) acetic acid for five minutes, rinsed with distilled H_2O and blotted dry. The gel was then transferred to Whatman 3MM paper, covered with 'Saran' wrap (Dow) and dried under vacuum on a Biowerk slab gel dryer (low heat, 2 hours). The dried gel was exposed to Fuji X-ray film (300 x 400mm)overnight at room temperature.

CHAPTER 3 DEVELOPMENT OF AN IN VITRO SYSTEM FOR THE TRANSCRIPTION AND TRANSLATION OF PLANT MITOCHONDRIAL DNA

3.1 INTRODUCTION

The first sections of this chapter describe attempts to develop a coupled *in vitro* transcription/translation system for the indentification and characterisation of protein coding genes in mt DNA. The reasons why these attempts using an *E. coli* system failed to respond to added plant mitochondrial nucleic acid together with possible ways of overcoming these problems are discussed in subsequent sections.

3.1.1 The potential use of *in vitro* systems for transcription and translation of DNA

The development of cell free systems capable of transcribing and translating DNA and translating RNA with fidelity has permitted the identification of the protein products of many genes. This approach has been particularly successful in the identification of protein coding genes in chloroplast genomes (reviewed in Whitfeld and Bottomley, 1983). Thus a DNA restriction fragment from chloroplast DNA (in a plasmid or as a linear DNA fragment) can be used as a template in the *in vitro* system, and the resulting ^{14}C or ^{35}S labelled polypeptides (if any) identified by coelectrophoresis with purified proteins, or immunoprecipitation with antibodies directed against specific chloroplast proteins. This approach has been used to identify and map several chloroplast genes from a variety of plants including the large subunit of RuBP carboxylase in spinach (Bottomley and Whitfeld 1979), maize (Bedbrook *et al* 1979) and *Chlamydomonas* (Malnoe *et al* 1979); and the *atp*B, *atp*E and *atp*H genes in wheat

(which encode subunits β and ε of CF_1 and subunit 9 of CF_o ATP synthase respectively (Howe et al 1982, Howe et al 1982a).

In contrast, protein coding genes in plant mitochondria have to date been identified by a number of less direct approaches. In yeast, identification of mt genes has been expedited by the large number of mitochondrial mutants available. Only two non lethal mutations in plant mt DNA ie sensitivity to T toxin/cytoplasmic male sterility (Leaver and Gray 1982) and non-chromosomal stripe (K. Newton, pers. commun.) have been identified. Also the plant mitochondrial genome is very large and is therefore not amenable to complete sequence analysis unlike mammalian mt DNA. The most direct available approach for the identification of plant mt protein coding genes has been by cross hybridisation under 'heterologous' conditions between plant mt DNA fragments and specific mitochondrial gene probes from yeast and mammals (see for example Fox and Leaver 1981, Dawson et al 1984, Isaac et al 1984). Final identification then requires DNA sequence analysis. The fact that these genes are actively expressed and are not silent copies is then confirmed by Northern blotting of mt RNA (Chapter 4, this thesis). This approach is clearly limited a) by the availability of DNA probes with sufficient homology to cross hybridise (for example animal and yeast probes for F_o ATPase subunit 9 consistently fail to hybridise to plant mt DNA (A. Dawson pers. commun.) even though the polypeptide is synthesised by isolated plant mitochondria (Hack and Leaver 1984), and b) by the possibility that plant mitochondria have a much higher coding capacity than other mt genomes and could therefore contain genes for which DNA probes of mitochondrialorigin are not available. Indeed, maize mt DNA has recently been shown to encode the F_1 ATPase α subunit which

is not found in mammalian and fungal mt genomes (A. Brennicke pers. commun.).

Another indirect approach used for the identification of protein coding genes has been to characterise the polypeptides which are synthesised by isolated mitochondria and are, therefore, by extrapolation, encoded in the mt DNA (see for example Hack and Leaver 1983, Hack and Leaver 1984). (On the assumption that RNAs are not imported into mitochondria)

The availability of an efficient in vitro transcription/translation system for mt DNA would therefore considerably improve our understanding of the organisation, information content and expression of the plant mt genome. Thus a restriction fragment of mt DNA could be added to a cell free transcription/translation system, and the resulting polypeptides (if any) analysed and compared with purified mt proteins (for example by coelectrophoresis, immunoprecipitation, partial proteolysis etc.). Using restriction enzymes selectively to digest the template DNA, the exact location of the gene on that fragment, and its polarity (ie. direction of transcription) can be determined. Hybridisation of the template DNA to a library of mt DNA clones would allow the location of that gene to be determined in the mt genome. A cell free translation system for mt RNA would also permit studies on the expression of mt genes, for example by measuring the levels of translatable mt mRNAs during plant development (Chapter 5).

3.1.2 Previous attempts at cell free transcription/translation of mitochondrial nucleic acids

Since rat liver mt DNA was shown to stimulate protein synthetic activity in an E. coli (Modollel 1971) cell free system (Rabussay et al 1969) there have been many attempts to obtain full length

translation products from yeast, mouse and rat mt DNA or RNA in a variety of heterologous cell free systems, most of which have been unsuccessful (see for example Chuang and Weissbach 1973, Chang et al 1975, Halbreich et al 1975). Yeast mt RNA fails to stimulate protein synthetic activity in cell free systems derived from reticulocytes and Xenopus oocytes (Moorman et al 1977), whilst only short length polypeptides (M_r <12,000) are produced in wheat germ (Roberts and Paterson 1973) and E. coli systems using mt DNA (Moorman and Grivell 1976, Scragg and Thomas 1975, Padmanaban et al 1975) or mt RNA (Moorman et al 1978) as templates. These low molecular weight polypeptides have been shown to contain the antigenic determinants for cytochrome oxidase subunits I and II (Moorman et al 1976). Up to 27% of the products synthesised in an E. coli cell free system under the direction of an 11 - 13S fraction of mt RNA are reactive with antisera raised against yeast cytochrome oxidase subunits I and II. Thus while the system was incapable of yielding full length translation products, it did suggest that these polypeptides were encoded in the yeast mt genome (Moorman et al 1978), since as yet there is no convincing evidence for import of RNA into mitochondria. Rat liver mt DNA stimulates an E. coli cell free system eight fold over endogenous activity, but again directs the synthesis of low molecular weight polypeptides which show no resemblance to the polypeptides synthesised by isolated mitochondria (Chuang and Weissbach 1973, Chang et al 1975).

The failure of mt RNA from a variety of organisms to be correctly translated in heterologous cell free systems is partly attributable to the altered genetic code used in mitochondria. In this context the most significant change is the use of UGA in human and yeast

mitochondrial genes to specify tryptophan rather than translation-termination (Barrell et al 1979, Fox 1979, Macino and Tzagoloff 1979). Thus, translation of yeast and mammalian mt RNAs in heterologous cell free systems will be terminated prematurely whenever an 'in frame' UGA^{trp} codon is encountered. In an attempt to overcome this problem, partially purified mRNA for yeast COII was translated in a wheat germ system supplemented with a suppressor tRNA ($tRNA^{Ser}_{UGA}$) from Schizosaccharomyces pombe. The inclusion of the suppressor tRNA led to the appearance of a high molecular weight polypeptide (M_r 31,000: wild type is M_r 29,000) among the translation products, which was immunoprecipitated with COII antibodies. RNA from two well defined mutants at this locus directed the synthesis of characteristically shortened polypeptides (De Ronde et al 1980). Since the mRNA for COII contains five UGA codons (Fox 1979) translation and suppression must have proceeded extremely efficiently.

In spite of the promise of such a modified system for the translation of mt RNAs, so far only yeast COII mRNAs have proved translatable in these systems with fidelity (L. Grivell pers. commun.). Failure of the system to translate other yeast mt mRNAs could be due to a variety of reasons, for example:

1) Many yeast mRNAs have long 5' leaders which may not be crossed by bacterial ribosomes. For example, F_o ATPase 9 mRNA has a leader of 500 nucleotides (Hensgens et al 1979). Exceptionally, the yeast COII mRNA does not have a long leader.

2) Lack of ribosome binding sites which could be recognised by E. coli ribosomes. This possibility is discussed later in section 3.3.

3). Several yeast mt genes contain introns, and hence only processed and spliced transcripts will yield an authentic protein product.

Two laboratories have described the reconstitution of homologous systems capable of using yeast mt RNA to direct the synthesis of a spectrum of polypeptides very similar to those synthesised *in vivo* and by isolated mitochondria (Kroon 1963, Pfisterer and Butow 1981). Such homologous systems are clearly less easy to prepare in any quantity, and also suffer the disadvantage that endogenous messages are difficult to eliminate. Several laboratories have since tried to repeat the reconstitution of these *in vitro* mt protein synthetic systems, but have so far failed (L. Grivell pers. commun.).

3.2 DEVELOPMENT OF A CELL FREE SYSTEM CAPABLE OF TRANSCRIBING AND TRANSLATING HIGHER PLANT MITOCHONDRIAL NUCLEIC ACIDS

3.2.1 Rationale and aims

The work described in this chapter was initiated against the background of repeated failure to develop a convenient *in vitro* system capable of translating yeast and mammalian mt RNAs with fidelity. There were several reasons to believe, however, that some of the problems encountered with the *in vitro* transcription/translation and translation of yeast and mammalian mt nucleic acids may not apply to plants. These were:

1) While the genetic code in plant mitochondria differs from both the universal code and that used by yeast and mammalian mitochondria [eg CGG (normally Arg) appears to encode Trp; Fox and Leaver 1981, A. Dawson, P. Isaac pers. commun.], there is no evidence that UGA (normally stop) encodes tryptophan. Indeed, UGA is thought to act as a stop codon in the *Oenothera COB* gene (A. Brennicke pers. commun.). Of the five protein coding genes (*COI*, *COII*, *COB*, F_1 ATPase α and subunit F_o ATPase subunit 8) sequenced from a variety of plants, no

in-frame UGA codons have been found. Hence there should be no major barrier preventing the *in vitro* translation of plant mt RNAs. Substitution of Arginine for Tryptophan in plant mt polypeptides by a heterologous cell free system which recognises CGG as an Arginine codon could lead to shifts in their apparent molecular weight but presumably would not significantly affect immunoprecipitation of the whole product.

2) Most plant mt genes sequenced to date do not contain introns whereas the homologous genes in yeast do. Although virus SV40 genes containing introns have been identified by cell free transcription/translation and immunoprecipitation (Roberts *et al* 1975) coupled cell free transcription/translation systems have been more successful in the identification of genes lacking introns. Of course split genes can be identified by hybrid release translation (Ricciardi *et al* 1979) or hybrid arrested translation (Paterson *et al* 1977), although this approach is less direct and does not provide such detailed information about the restriction map position and the direction of transcription of the gene.

3.2.2 Approach

The approach taken to the development of a cell free transcription/translation system for higher plant mt nucleic acids was as follows:

1) Choose an existing standard cell free system for development and modification

2) Ascertain the cell free system chosen could transcribe and translate heterologous eg viral DNA and translate heterologous eg chloroplast mRNA with fidelity

3) Obtain and optimise translation of mt RNA into full length polypeptides. These polypeptides would be identified a) by

immunoprecipitation with heterologous (for example yeast) or homologous antibodies or b) where no antibodies exist, coelectrophoresis, partial proteolysis and two dimensional peptide mapping in comparison with purified polypeptides.

4) Optimise the system for linked transcription/translation of restriction fragments of plant mt DNA.

5) Examine the translation products synthesised under the direction of mt RNA in a different translation system (for example wheat germ or reticulocyte lysate), since it has been shown, for example, that some chloroplast RNAs are not translated in an E. coli cell free system but are translated in a reticulocyte system (Coen 1982).

The E. coli cell free system (Modollel 1971) was chosen for initial studies for the following reasons:

1) There is circumstantial evidence that plant mitochondria (together with yeast and human mitochondria) initiate polypeptide chains with N-formyl methionine, in common with prokaryotes but not eukaryotes (Anderson et al 1977). Phaseolus vulgaris mitochondria contain three methionyltRNAs, two of which can be formylated by a mitochondrial or E. coli transformylase enzyme (Guillemaut and Weil 1975). Furthermore, Gray and Spencer (1983) have shown by sequence analysis that wheat mt DNA encodes a eubacteria-like initiator methionine $tRNA_f^{met}$.

2) Higher plant mitochondrial ribosomes, although larger than bacterial ribosomes (78S cf 70S, Leaver and Harmey 1973) are sensitive to the same major groups of antibiotics which inhibit eubacterial protein synthesis, for example chloramphenicol (Clark-Walker and Linnane 1976), and are insensitive to inhibitors of cytosolic ribosomes, for example cycloheximide (Lamb et al 1968, Borst and Grivell 1971).

3.2.3 Translation of heterologous RNA by the E. coli cell free system

An essential factor in the development of a cell free system to translate mt RNA was to demonstrate that the in vitro system could translate added exogenous non-mitochondrial mRNA with fidelity. Cucumber cotyledon total RNA from five day old light grown seedlings was used as a template. The conditions for incorporation were those which had previously been determined to be optimal for the incorporation of ^{35}S methionine into protein/acid insoluble material under the direction of chloroplast RNA (Walden and Leaver 1981).

Table 3.1 shows the response of the E. coli in vitro system to added cucumber RNA, incubated under the conditions described in Chapter 2.

Table 3.1 Stimulation by added cucumber total cellular RNA of protein synthesis in an E. coli in vitro system

Sample	Incorporation into protein (cpm ^{35}S) per incubation	fold stimulation over endogenous activity
Zero time	17,400	-
Endogenous RNA	27,700	-
+ cucumber total RNA (15µg) (from five day old, light grown plants)	209,000	

Mixtures were incubated for 20 minutes at 37°C.

The translation products synthesised in the E. coli cell free system under the direction of cucumber total RNA were analysed by electro-

phoresis through a 15% (w/v) polyacrylamide/0.1% (w/v) SDS gel. The molecular weights of the translation products were estimated using a calibration curve of $\log_{10}$ molecular weight of standard protein markers against mobility (Weber and Osborn 1969). Figure 3.1 shows the translation products of cucumber RNA-stimulated protein synthesis. Discrete polypeptides in the range M_r 70,000 - <10,000 are synthesised. Particularly prominent is a polypeptide with a molecular weight of approximately 55,000. It is assumed that this represents the chloroplast genome-encoded (Whitfeld and Bottomley 1983) large subunit of ribulose 1.5 bisphosphate carboxylase previously shown to be translated with fidelity in an E. coli cell free system and to have a molecular weight of ca 55,000 (Hartley et al 1975, Bottomley et al 1976). Of course criteria other than molecular weight are required to unequivocally identify an in vitro synthesised polypeptide since such polypeptides often differ in molecular weight to their in vivo counterparts (Coen 1982, Langridge 1981). However, for the purpose of the experiments described here, electrophoretic mobility of the translation products was the main criterion by which the fidelity of translation was assessed.

If the M_r 55,000 polypeptide is the RuBPCase large subunit, then the finding that it is a major product of the E. coli system when programmed with cucumber total cellular RNA is in agreement with the work of Bottomley et al (1976) and of Walden and Leaver (1981) who found that the spectrum of polypeptides synthesised in an E. coli in vitro system under the direction of chloroplast RNA and total spinach RNA were identical. In particular the mRNA encoding the large subunit of RuBPCase from many plants seems to be a good template for translation in the E. coli system (Bottomley and Whitfeld 1979) and is

Fig. 3.1 Translation products of the E. coli cell free translation system directed by cucumber total cellular RNA

The E. coli cell free translation system was incubated with 15µg RNA isolated from five day old light grown cucumber plants. The polypeptides synthesised in the cell free system were analysed by electrophoresis through a 15% (w/v) polyacrylamide/0.1% (w/v) SDS gel followed by autoradiography.

3.1 b Autoradiogram of ^{35}S-labelled polypeptides synthesised by the E. coli cell free system in the absence of added RNA ie endogenous mRNA control.

3.1 a Autoradiogram of ^{35}S-labelled polypeptides synthesised by the E. coli cell free system under the direction of 15µg cucumber RNA.

Equal proportions of the S-30 were loaded on each track.
Molecular weight markers were standard proteins of known molecular weight which were visualised by staining with Coomassie brilliant blue.

Fig. 3.2 Translation products of the E. coli cell free transcription/ translation directed by bacteriophage T7 DNA

The E. coli cell free translation system was incubated with 2.5µg T7 DNA. The polypeptides synthesised in the cell free system were analysed by electrophoresis through a 15% (w/v) polyacrylamide/0.1% (w/v) SDS gel followed by autoradiography.

3.2 a Autoradiogram of ^{35}S-labelled polypeptides synthesised by the E. coli cell free system in the absence of added DNA ie endogenous DNA/mRNA control.

3.2 b Autoradiogram of ^{35}S-labelled polypeptides synthesised by the E. coli cell free system under the direction of 2.5µg T7 DNA.

Equal proportions of the S-30 were loaded on each track.
Molecular weight markers were visualised as described for Fig. 3.1.

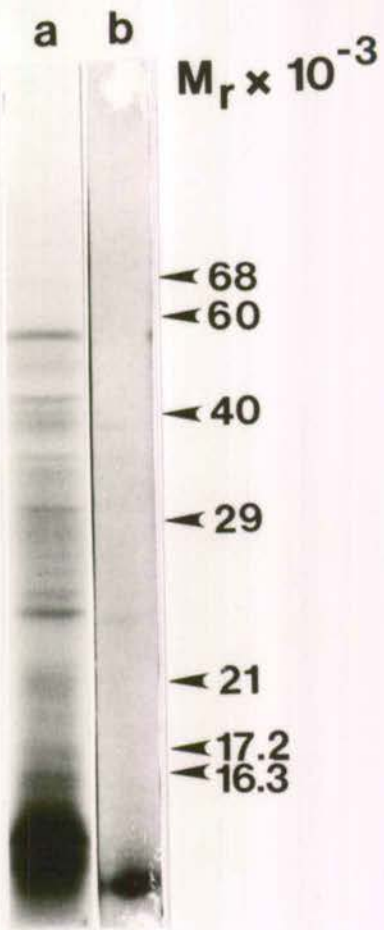

a
b
$M_r \times 10^{-3}$
68
60
40
29
21
17.2
16.3

Fig. 3.1

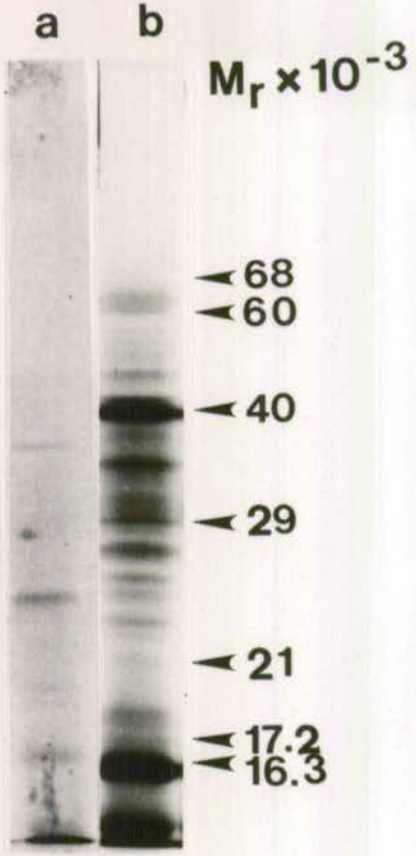

a
b
$M_r \times 10^{-3}$
68
60
40
29
21
17.2
16.3

Fig 3.2

frequently the most prominent product from such translation systems.

3.2.4 Transcription/translation of heterologous DNA in the E. coli cell free system

In order to determine whether the E. coli system could correctly transcribe and translate exogenous DNA, bacteriophage T7 DNA, which encodes about 42 well characterised proteins (Studier and Rosenberg 1981) was used to programme the system. The E. coli translation system (section 3.2.3) was modified for coupled transcription/translation as described in materials and methods, using essentially the conditions previously optimised for the transcription/translation of chloroplast DNA (Bottomley and Whitfeld 1979).

The bacteriophage T7 genome comprises a double-stranded linear DNA of 40 Kb (Dubin et al 1970). The polypeptides it encodes are classed (classes I, II and III) according to the temporal order of their expression in the host, E. coli [see review by Studier (1972)]. Upon infection, class I genes are transcribed by the host RNA polymerase. Class II and III genes are then transcribed by a 'phage encoded RNA polymerase (itself a class I gene). Normally transcription by the host RNA polymerase beyond the class I gene region is prevented by a transcription termination signal. However, termination is not completely effective and class II and III genes can be transcribed by the host RNA polymerase.

Bacteriophage T7 DNA (2.5μg) was found to stimulate [^{35}S] methionine incorporation into acid insoluble material/protein in an E. coli system by approximately fifteen-fold (Table 3.2).

Table 3.2 Stimulation by added bacteriophage T7 DNA of protein synthesis in an E. coli in vitro system

Sample	Incorporation into protein (cpm) per incubation	fold stimulation over endogenous activity
Zero time	35,000	-
Endogenous RNA (no additional UTP or CTP)	63,000	-
Endogenous DNA (+ UTP and CTP)	91,000	-
+ T7 DNA (2.5µg)	1,402,760	24

Mixtures were incubated for 20 minutes at 37°C.

Figure 3.2 shows the translation products synthesised by the E. coli cell free system under the direction of T7 DNA. The polypeptides were characterised on the basis of molecular weight only, as discussed in section 3.2.3 . Particularly prominent are the gene products presumed to be Kinase (M_r 41,000), ligase (M_r 41,000), head assembly protein (M_r 40,000) and lysozyme (M_r 16,800) (Studier and Rosenberg 1981). The first three co-migrate as a major radioactive band. No prominent polypeptides are detected in the absence of added DNA. These results confirm those of Moorman et al (1976) who found that T7 class II genes (for example lysozyme) and class III genes are transcribed and translated in an E. coli cell free system. A simple check on the fidelity of transcription and translation could be made by programming the system with 'phage T7 DNA isolated from an amber mutant which should direct the synthesis of shortened polypeptide(s)

in the absence of a suppressor tRNA.

It appears, therefore, that the E. coli system is able to transcribe and translate exogenous DNAs and translate heterologous RNAs with fidelity when the products are analysed on the basis of molecular weight only.

3.2.5 Optimising the E. coli system for the translation of mt RNA

Following the demonstration that the E. coli in vitro system could transcribe and translate heterologous DNAs and RNAs with fidelity, attempts were made to optimise the system for the translation of plant mt RNA. The in vitro system was initially optimised for translation rather than coupled transcription/translation for two reasons:

1) At the time these experiments were initiated the only cloned and sequenced protein coding plant mitochondrial gene was that encoding COII from maize (Fox and Leaver 1981). This gene is interrupted by a 794 bp intron. Thus if the DNA was correctly transcribed by E. coli RNA polymerase, the resulting RNA would contain unspliced intron sequences. The intron itself contains a TGA in frame with the first exon, and so translation of the unspliced RNA would lead to synthesis of a shortened COII polypeptide containing 33 intron-encoded amino acids fused with the exon 1 encoded amino acids. Since the only available antibodies for COII were those from yeast (shown to cross react with the maize COII polypeptide), it was not known if such heterologous antibodies would cross react with such a shortened COII polypeptide (although they almost certainly would).

2) The major barrier to the in vitro expression of yeast and

mammalian mt genes appears to be translation rather than transcription especially because of the occurrence of 'in-frame' UGA codons specifying tryptophan in mitochondria but 'stop' in E. coli (De Ronde et al 1980, Moorman et al 1976). Therefore, if plant mt RNA can be reproducibly translated in vitro and with fidelity, in theory it should be relatively easy to couple the system for both transcription and translation.

The maize mt RNA used in the following experiments was isolated using the detergent NDS (Chapter 4, section 4.2.2.1) and thus contained low levels of DNA (Chapter 4, section 4.2.2.1). Mt nucleic acids isolated in this way were not purified further so the bulk of the nucleic acid added was ribosomal and transfer RNA.

3.2.5.1 Effect of maize mt RNA concentration on incorporation of ^{35}S L-methionine in the E. coli cell free system

To investigate the response of the E. coli cell free system [essentially as optimised for the translation of chloroplast RNA (Bottomley and Whitfeld 1979, Walden and Leaver 1981)]the system was primed with increasing amounts of maize mt RNA. Figure 3.3 shows that the E. coli cell free system is very rapidly saturated by mt RNA. Thus addition of more than 5µg mt RNA per incubation does not lead to additional incorporation of ^{35}S-methionine into protein. It is possible that the mt RNA preparation contained some contaminant that inhibited translation by E. coli ribosomes. This possibility could have been investigated by mixing for example chloroplast and mitochondrial RNAs, and determining the level of translation achieved compared to chloroplast RNA mixed with an equivalent amount non-translatable RNA (eg chloroplast rRNA). Subsequent investigations

Fig. 3.3 Response of the E. coli S30 to maize mt RNA

Increasing amounts of maize mt RNA (0 - 30µg) were used to programme the E. coli in vitro translation system. Incorporation of ^{35}S methionine into protein/acid insoluble material was estimated by spotting duplicate 5µl aliquots of the reaction mixture onto squares of Whatman 3MM paper and estimating radioactivity as described in Chapter 2.

Fig. 3.4 Translation products of the E. coli cell free translation system directed by maize mt RNA

The E. coli cell free system was incubated with 5ug maize mt RNA. The polypeptides synthesised in the cell free system were analysed by electrophoresis through a 15% (w/v) polyacrylamide/0.1% (w/v) SDS gel followed by autoradiography.

3.4 a Autoradiogram of ^{35}S labelled polypeptides synthesised by the E. coli system in the absence of added RNA ie endogenous RNA control.

3.4 b Autoradiogram of ^{35}S labelled polypeptides synthesised by the E. coli cell free system under the direction of 5µg mt RNA.

Equal proportions of the S30 were loaded on each track.

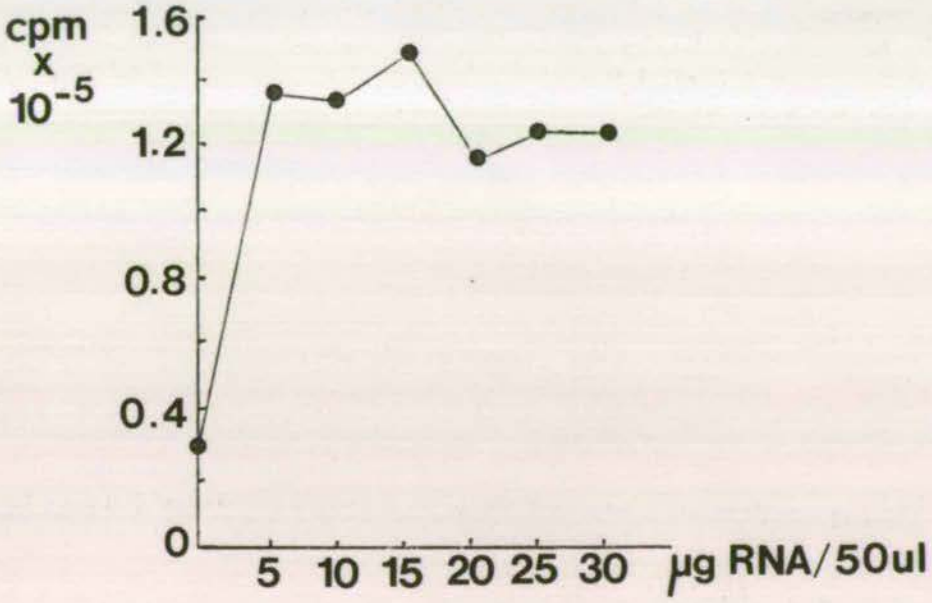

cpm
x
10-5
1.6
1.2
0.8
0.4
0
5
10
15
20
25
30
µg RNA/50ul

Fig. 3.3

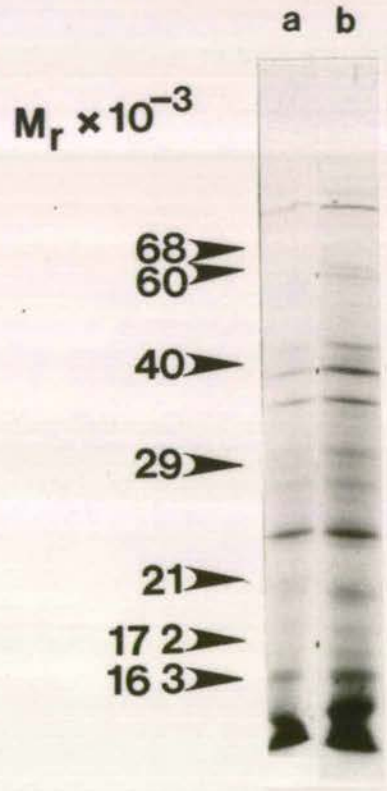

a
b
Mr x10-3
68
60
40
29
21
17 2
16 3

Fig. 3.4

(A. Liddell pers. commun.) have confirmed that the E. coli system is indeed saturated with 5 - 10µg RNA/50µl incubation. Mt RNA (5µg) stimulates ^{35}S methionine incorporation into protein approximately four-fold over the endogenous control activity. Fig. 3.4 shows the ^{35}S-labelled translation products synthesised under the direction of 5µg maize mt RNA. Although the endogenous RNA control track is fainter it is clear that mt RNA does not direct the synthesis of any novel polypeptides compared to the endogenous RNA when analysed by autoradiography. It is likely therefore that novel polypeptides, if any, are present at such a low level that they could only be detected by immunoprecipitation.

3.2.5.2 Effect of $[Mg^{2+}]$ and $[K^{+}]$ on translation of mt RNA in the E. coli cell free system

Initial translation experiments with mt RNA were carried out under the conditions optimised for the in vitro translation of chloroplast RNA (Walden and Leaver 1981). Since the concentrations of Mg^{2+} and K^{+} can have a significant effect on both the quantity and the quality of polypeptides synthesised in the E. coli cell free system (Bottomley 1982), the effect of altering their concentration between 7.5 and 15mM on the in vitro translation of mt RNA was investigated. The optima for Mg^{2+} and K^{+} in terms of maximal incorporation of ^{35}S methionine into protein were determined to be ca 11mM and 10mM respectively. However it is important to note that endogenous activity was not measured over the range of K^{+} and Mg^{2+} concentrations examined. This factor could be important if mt RNA is having a 'sparing' effect on endogenous E. coli RNA (see below and Chuang and Weissbach 1973). Even at the optimal Mg^{+} and K^{+} concentrations, however, incorporation of ^{35}S methionine into protein is still very low, and at most 4 - 6-fold over endogenous activity.

3.3 DISCUSSION

It is clear from these very preliminary experiments that translation of mt RNA and transcription/translation of mt DNA will not be achieved in a conventional cell free system with minor modifications. The major problem in these experiments has been the low level of stimulation effected by the addition of mt RNA, and the resulting comparatively high level of endogenous translation activity in the system. This hampers the detection of novel polypeptides synthesised (if any) under the direction of mt RNA. It is possible that the addition of mt RNA in some way protects the endogenous E. coli RNA. Hence the stimulation of activity seen when mt RNA is added to the system may simply be due to translation of endogenous RNA at a greater rate. One way to examine this would be to look at the ratios of two labelled amino acids incorporated in the presence and absence of mt RNA. Thus for example the E. coli cell free system could be incubated with two amino acids, one labelled with ^{14}C and the other with ^{3}H. Determination of the ^{14}C:^{3}H ratios in the polypeptides synthesised (a) in the absence and (b) in the presence of mt RNA would determine whether mt RNA was indeed 'sparing' the endogenous RNA.

Since it is now thought that plant mt mRNAs do not contain in-frame TGA codons, it is surprising that the E. coli system shows so little response to added plant mt RNA. This, coupled with the fact that only one yeast mt mRNA has been translated with fidelity in a cell free system suggests that some even more fundamental barrier to translation exists.

Specificity of translation initiation in bacteria is thought to involve base pairing of a sequence between three and nine nucleotides

long situated between three and eleven nucleotides from the AUG initiation codon to a complementary sequence near the 3' terminus of the 16S rRNA (Shine and Dalgarno 1974, Steitz and Jakes 1975). This oligonucleotide sequence is now known as the Shine and Dalgarno sequence.

Analysis of sequences at the 3' ends of the wheat mt 18S (Spencer et al 1984) and maize mt 18S (Chao et al 1983) rRNAs has shown that while the overall sequence of the 18S rRNA is highly homologous to the E. coli 16S rRNA sequence, part of the Shine and Dalgarno sequence (5' CACCUCCU 3') is replaced by a novel sequence 5' UGAAUCCU 3' in both wheat and maize mitochondria (post transcription processing removes part of this sequence which is therefore either 5' UGAAUCC 3' or 5' UGAAUC 3'). Analysis of known plant mt genes has revealed that in most cases a sequence occurs approximately 15 to 20 nucleotides 5' of the AUG codon which shows 3/4 or 4/4 complementarity with the higher plant-specific sequence (5' UGAA 3') in the mt rRNA Shine and Dalgarno region. In most cases additional homology to the remainder of the sequence occurs (Dawson et al 1984). The functional significance of these sequences, if any, in the initiation of translation of plant mt mRNAs remains to be established, for example by ribosome protection experiments (Ravetch et al 1977).

If this sequence does have functional significance in the formation of a translation/initiation complex in higher plant mitochondria the low stimulation of incorporation by mt RNA in the E. coli cell free system is not unexpected. E. coli ribosomes would not bind so efficiently to the corresponding sequence in the plant mitochondrial

mRNA. The importance of E. coli like ribosome binding sites in translation of heterologous mRNAs by the E. coli cell free system is not fully understood. Some chloroplast mRNAs eg LS RuBPCase from spinach (Zurawksi et al 1981), pea (Zurawski et al 1983) and maize (McIntosh et al 1980) have eubacterial-like Shine and Dalgarno sequences 5' of the AUG initiation codon. However, no such sequences capable of base pairing with the E. coli 16S rRNA are found 5' to the genes encoding atpB in spinach (Zurawski et al 1982) or pea (Zurawski et al 1983) and yet these mRNAs are translated in an E. coli cell free system, although at somewhat lower levels than LS RuBPCase (Whitfeld and Bottomley 1983). Thus the high efficiency with which the mRNA for LS RuBPCase is translated by the E. coli cell free system may be attributable to the presence of a sequence that promotes translation initiation.

It is possible that once a suitable initiation complex had formed with plant mt mRNAs, translation could proceed normally in an E. coli cell free system, in view of the seemingly eubacterial fashion of protein synthesis in plant mitochondria. Hence an E. coli cell free system supplemented with plant mt ribosomes might be capable of in vitro translation of plant mt mRNAs.

The lack of success in the development of a cell free system capable of translating mt RNAs (let alone transcribing and translating mt DNA) has greatly hindered identification of the genes encoding the 20 - 50 polypeptides synthesised by isolated mitochondria. Whether supplementation of an E. coli cell free system with mitochondrial ribosomes and other components will overcome this problem, or whether some even more fundamental barrier exists to the translation of mt RNAs in

heterologous systems, will only be resolved further by detailed experimentation.

CHAPTER 4 CHARACTERISATION OF THE TRANSCRIPTS OF THE MAIZE MITOCHONDRIAL GENOME

4.1 INTRODUCTION

This chapter describes the characterisation of the transcripts of the maize mt genome, particularly those of the genes encoding cytochrome oxidase subunits I & II (COI and COII respectively) apocytochrome b (COB) and F_1 ATPase α subunit. The first sections of this chapter deal with the optimisation of techniques for studying mt gene expression in maize. The results of these analyses are presented and discussed in subsequent sections.

4.1.1 Aims and rationale

Until this work was initiated, the main emphasis in the study of the plant mt genome had been to identify, isolate and sequence mitochondrial genes (Fox and Leaver 1981, Dawson et al 1984, Isaac et al 1984, Hiesel and Brennicke 1983, Dale et al 1984). Apart from the identification and sequencing of rRNAs (Leaver and Harmey 1973, Pring and Thornbury 1975, Spencer et al 1981) transcription of the plant mt genome had barely been investigated. The aims of the work presented in this chapter were as follows:

1) To demonstrate that the protein coding genes identified in mt DNA by sequence analysis are actively expressed. In the absence of a coupled transcription/translation system for mt DNA (Chapter 3) the best means to identify active mt genes is by nucleotide sequencing and transcript analysis. It is important to show that genes tentatively identified by sequence analysis are actively expressed, since there are examples of mt genomes harbouring silent copies of genes (van den Boogaart et al 1982, Scazzocchio et al 1983).

2) A full understanding of the expression of the plant mt genome requires that the transcripts of each gene are characterised in detail including, for example: their length in relation to the protein coding sequence, identification of polycistronic messages, identification of promoter sequences and associated control regions, and the elucidation of processing pathways of primary transcripts.

For these reasons detailed transcript analysis of four mt protein coding genes (COI, COII, COB, α subunit ATPase) was undertaken. These would represent the first plant mitochondrial protein encoding RNAs to be unequivocally identified and characterised in detail, thus making a significant contribution to our understanding of the expression of the plant mitochondrial genome.

4.2 APPROACH

Of the methods available for studying the transcription of the maize mt genome, the techniques of "Northern" blotting (Alwine et al 1977) and nuclease S1 mapping (Berk and Sharp 1977) were chosen for the majority of work described in this chapter. Northern blotting involves fractionation of mt RNAs by electrophoresis through denaturing agarose gels, transfer of RNA to a support membrane followed by hybridisation with gene-specific DNA probes. This technique can provide preliminary information about the expression of particular genes, for example, transcripts can be accurately sized, and, by using probes from specific regions of mt DNA, mapped on the genome. Where transcripts undergo excision of more than one intron, splicing pathways can be tentatively elucidated (see for example Corruzzi et al 1981) although pulse chase experiments are required to determine unequivocally product-precursor relationships.

The major disadvantages of this technique are:

1) Only steady state levels of a transcript are measured - rate of synthesis and turnover cannot be estimated by this approach. Thus the technique can be used to differentiate only between regions of the genome which are transcribed strongly and/or whose transcripts are stable, and those which are weakly transcribed and/or whose transcripts are unstable.

2) The technique does not differentiate between translatable and non translatable mRNAs (see Gabrielli and Baglioni 1977, Leaver and Lovett 1974 for examples of masked mRNAs) and where more than one transcript hybridises to a particular DNA fragment, other approaches must be used to determine which is the mature mRNA, for example by *in vitro* translation (Ellis 1976).

Nuclease S1 mapping was used to map the locations of the 5' ends of the specific gene transcripts identified by Northern blotting.

4.2.1 Isolation of RNA

In order to characterise the *in vivo* transcripts of the maize mt genome, and to avoid artefacts, the following points must be considered:

4.2.1.1 Integrity of RNA

RNA must be intact and undegraded. This is particularly important if more than one transcript arises from a gene since it is difficult to distinguish between multiple transcripts and discrete breakdown products. Stern and Newton (1984) have recently reported that isolation of plant mt RNA by the detergent phenol method (Parish and Kirby 1966) leads to considerable breakdown unless the nuclease inhibitor aurintricarboxylic acid (ATA) is included in the extraction buffer.

Fig.4.1

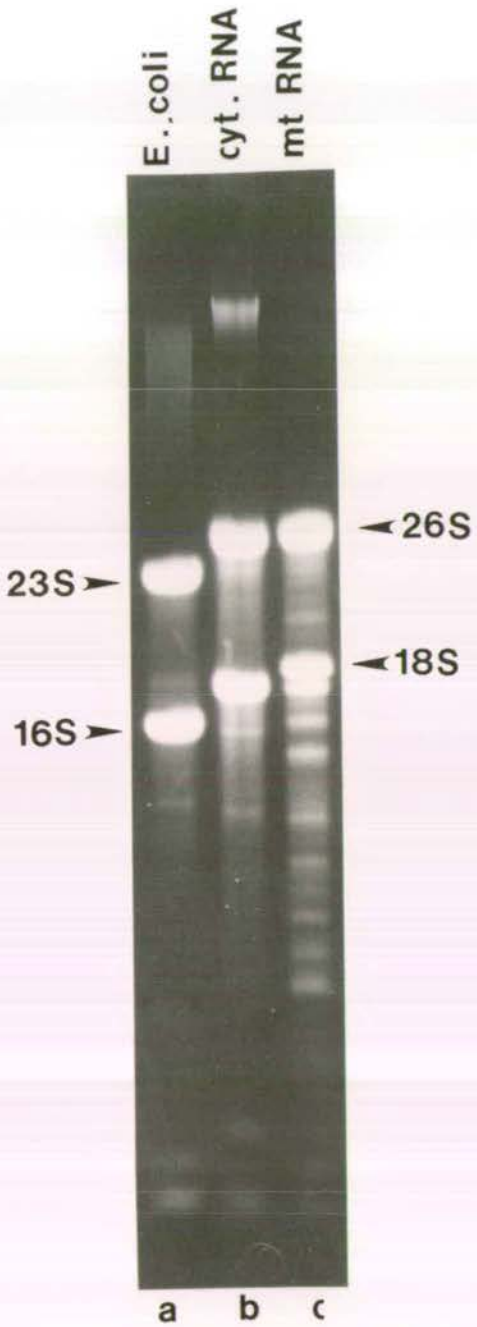


Fig 4.1 Gel electrophoresis of maize mitochondrial RNA

Maize mitochondrial RNA (10 µg isolated from four day old etiolated shoots) was denatured with urea and formamide and electrophoresed through a 2% agarose gel containing 10 mM NaPi pH 7.0 buffer. Total cellular RNA and E.coli RNA which had been treated identically were electrophoresed in parallel tracks to provide markers. RNA was visualised by staining with ethidium bromide and viewing under incident U.V. light.

Track a) 10 µg E.coli RNA

Track b) 10 µg total cellular RNA.

c) 10 µg maize mitochondrial RNA.
Note the multiplicity of ethidium bromide staining bands in this track.

This conclusion is reached from the pattern of RNA when analysed by electrophoresis through agarose gels and stained with ethidium bromide. RNA isolated in the absence of ATA gives rise to a large number of ethidium bromide staining bands whereas RNA isolated in the presence of ATA does not. In the work reported in this thesis, maize mt RNA was isolated without ATA and when electrophoresed does give rise to a large number of bands (Fig. 4.1). This multiplicity of bands is found with several different gel systems. Such complex RNA gel profiles are also a feature of yeast mitochondria (eg. Hensgens et al 1983, van Ommen et al 1979). Whether these multiple bands are native mt RNAs or represent discrete breakdown products is unresolved, since it is impossible to make any objective comparison with regard to the effect of ATA from the data presented by Stern and Newton (1984), because the RNA isolated in the absence of ATA is clearly overloaded on the gel compared to RNA isolated with ATA. However, mt RNA isolated in the experiments described in this chapter, in the absence of ATA, was intact and not significantly degraded as discussed below. Three separate approaches were used to measure the integrity of mt RNA.

4.2.1.1.1 Sucrose gradient centrifugation of RNA A useful indicator of RNA breakdown is to examine the absorbance ratio of large ribosomal to small ribosomal RNAs (Leaver and Harmey 1973). Mt RNA was layered onto a linear 15 - 30% (w/v) sucrose gradient (McConkey 1967) in 0.1M Tris, pH 7.5, 0.5M NaCl, 5mM EDTA, 0.5% (w/v) SDS and centrifuged at 29,000 rpm for 18 hours in a 6 x 14 MSE swing out rotor at 20°C. The gradient was displaced from the top by injection of a 60% (w/v) sucrose solution. The gradient was passed

through an ultraviolet analyser (Instrumentation Specialities Co. Inc., Nebraska) and optical density (A_{260}) recorded continuously. Fig. 4.2 shows the separation profile of 26S and 18S rRNAs. The ratio of these RNAs is 1.5:1 which is less than the value of 1.8:1 expected from a ribosome containing equimolar proportions (Leaver and Harmey 1973) of rRNA of molecular weights ca 3546 (26S, Dale et al 1984) and ca 1955 (Spencer et al 1984). This is most probably explained by slight degradation of the 26S rRNA, as shown by Leaver and Harmey (1973) and Pring and Thornbury (1975). The resolving power of the gradient was insufficient to estimate the amount of contaminating cytoplasmic 5.8S rRNA (Dyer and Leaver 1981).

4.2.1.1.2 Heterologous rDNA hybridisation Mt RNA was isolated from four day old etiolated maize seedlings, electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, blotted to nitrocellulose and hybridised with nick-translated E. coli total DNA under conditions described in section 2.2.4.5. E. coli rRNAs are known to hybridise strongly to plant mt DNAs (A. Brennicke pers. commun.) due to the high nucleotide conservation (overall homology of 76% between wheat mt 18S and E. coli 16S rRNAs, Spencer et al 1984; many regions of $>70\%$ homology between maize mt 26S and E. coli 23S; Dale et al 1984). Since seven copies of the ribosomal cistron are present in each E. coli genome (Kiss et al 1977) it was reasoned that nick-translated E. coli total genomic DNA would hybridise to mt rRNA (since mt rDNA probes were not available). ^{32}P labelled E. coli DNA was kindly supplied by J. Manson, Department of Molecular Biology, Edinburgh University. Fig. 4.3 shows that E. coli genomic DNA hybridises to E. coli rRNAs and to both 26S and 18S mt RNAs. Since only these two bands hybridise it was

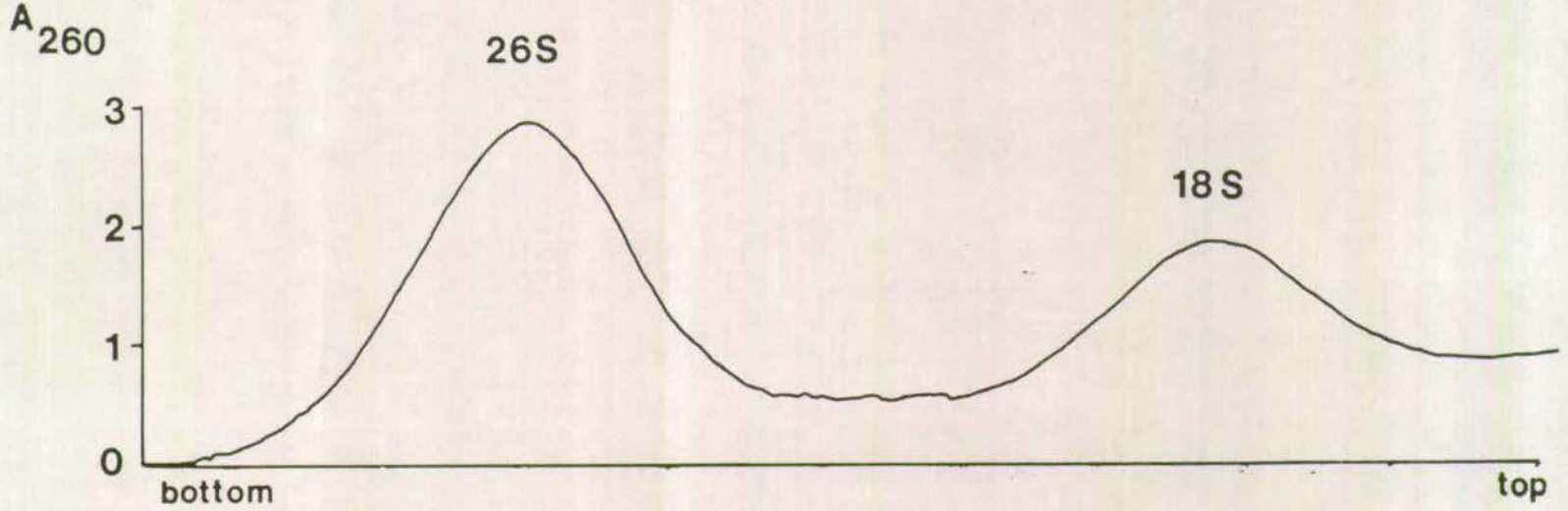


Fig 4.2 Sucrose gradient centrifugation of maize mitochondrial RNA

Maize mitochondrial RNA was layered onto a linear 15-30% (w/v) sucrose gradient in 0.1M Tris Cl pH7.5, 0.5M NaCl, 5 mM EDTA, 0.5% (w/v) SDS and centrifuged at 29,000 rpm for 18 hours in a 6 x 14 MSE swing out rotor at 20°C. The gradient was displaced from the bottom and passed through an ultraviolet analyser and optical density (A_{260}) continuously recorded. The ratio of 26S rRNA to 18S rRNA is 1.5:1.

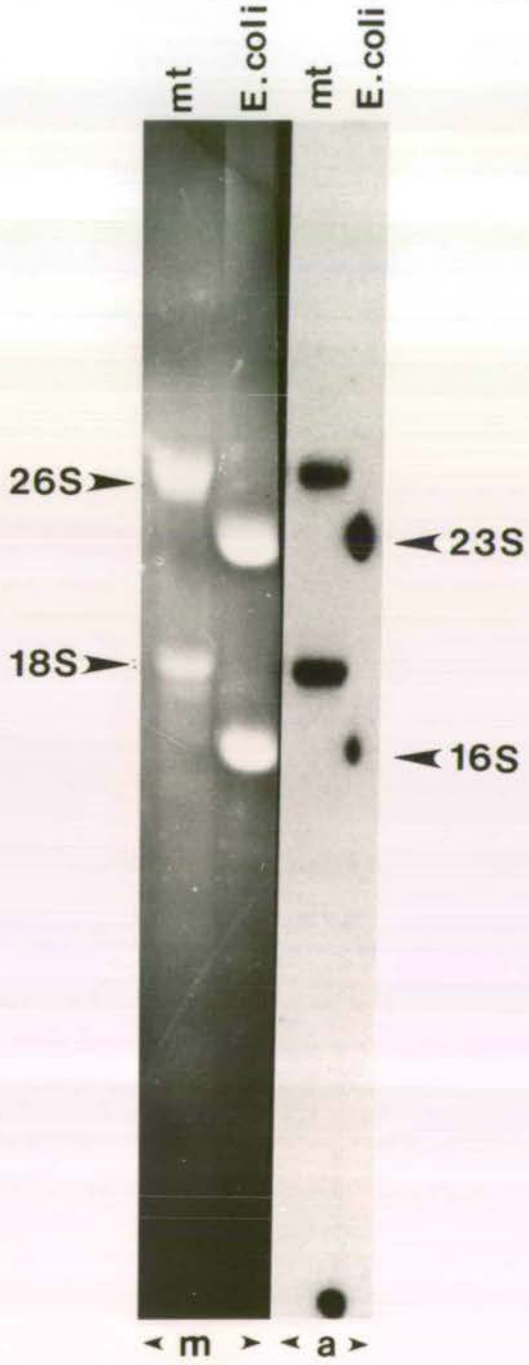


Fig. 4.3

Fig 4.3 Hybridisation of nick translated E.coli genomic DNA to maize mitochondrial RNA

Maize mt RNA (10 µg) and E.coli RNA (10 µg) were electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, blotted to nitrocellulose and hybridised with E.coli ^{32}P labelled total genomic DNA.

Tracks m) marker tracks of maize mt RNA and E.coli RNA stained with ethidium bromide and visualised under U.V. light. Tracks a) autoradiograms of E.coli ^{32}P labelled genomic DNA hybridised with maize mt and E.coli RNA.

concluded that the maize 18S and 26S rRNAs and the E. coli 23S and 16S rRNAs are intact and not significantly degraded. Unexpectedly, no 4S or 5S mt RNA was found to hybridise to the E. coli DNA. This is probably partly due to:

a) lower nucleotide sequence homology between E. coli and mt 5S rRNA (45% between wheat mt and E. coli 5S rRNAs, Spencer et al 1981) and

b) poor retention of 4S and 5S RNAs on nitrocellulose of pore size 0.45μm (Thomas 1983). The hybridisation of E. coli DNA to only two rRNAs in mt RNA also indicates that plastid contamination is insignificant, since any chloroplast RNAs, which would be resolved from the mt rRNA in the gel system used would also hybridise to E. coli DNA.

4.2.1.1.3 Comparison of mitochondrial transcripts in Northern blots of mt RNA and total cellular RNA The patterns of transcripts obtained when Northern blots of mt RNA and total cellular RNA were probed with defined mt DNA probes were compared with the following aims:

1) Damage of mitochondria during isolation could result in partial degradation of mt RNA. In contrast, isolation of total cellular RNA from maize shoots is very rapid, allowing less opportunity for nuclease action. Also, if in the absence of nuclear encoded enzymes transcription and/or processing is arrested, precursor/product ratios could change and so lead to artefacts when studying Northern hybridisation patterns in isolated mt RNA.

2) The contribution of any contaminating plastid and cytoplasmic RNA to the overall pattern of transcripts can be estimated by comparing the pattern of transcripts hybridising to defined gene probes in mt RNA and total cytoplasmic RNA. Any changes in the pattern of

bands, or the relative stoichiometry of bands, would indicate a contribution from contaminating RNAs. This was considered important since (i) two chloroplast proteins (cytochrome $\underline{b}_6$ and a 17K polypeptide; Widger *et al* 1983, Hauska *et al* 1983) show homology to maize apocytochrome *b* (although the regions of exact homology are very short, A. Dawson pers. commun.) (ii) there is increasing evidence that many mt genes share homology with sequences in the nucleus (Chapter 1, section 1.10) which if transcribed could hybridise to mt gene probes. A comparison of the pattern of transcripts in maize mt RNA and total cytoplasmic RNA hybridising to gene specific DNA probes for maize mt *COB* and *COI* (not shown) genes revealed essentially no differences. (Fig. 4.4)

From these investigations it was concluded that the mt RNA isolated from etiolated maize shoots and used in the following experiments was not significantly degraded or contaminated with plastid RNA.

4.2.2 Purity of mitochondrial RNA

4.2.2.1 Removal of DNA from mitochondrial RNA preparations

There is increasing evidence that the plant mt genome can undergo extensive recombination across repeated elements to form small circular molecules, some of which may fall into the size class range of mt transcripts (Palmer and Shields 1984, Lonsdale *et al* 1983). Since these recombinant molecules may contain copies or partial copies of genes it was considered important to remove DNA from RNA preparations. Inclusion of the detergent NDS in mt RNA isolation was found to considerably reduce the amount of mt DNA in preparations of mt RNA (Kirby 1965). Digestion of the sample with DNase I (RNase free) was sufficient to remove remaining DNA as judged by gel electrophoresis. (Fig. 4.5).

Fig 4.4 Hybridisation of maize COB ^{32}P labelled internal DNA probe to maize mt RNA and total cellular RNA

Maize mt RNA (5 µg) and total cellular RNA (20 µg) were electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with a DNA probe internal to the protein coding sequence (M13 clone 979, defined section 4.5.2).

Track a) maize mt RNA

Track b) total cellular RNA.

Fig. 4.4

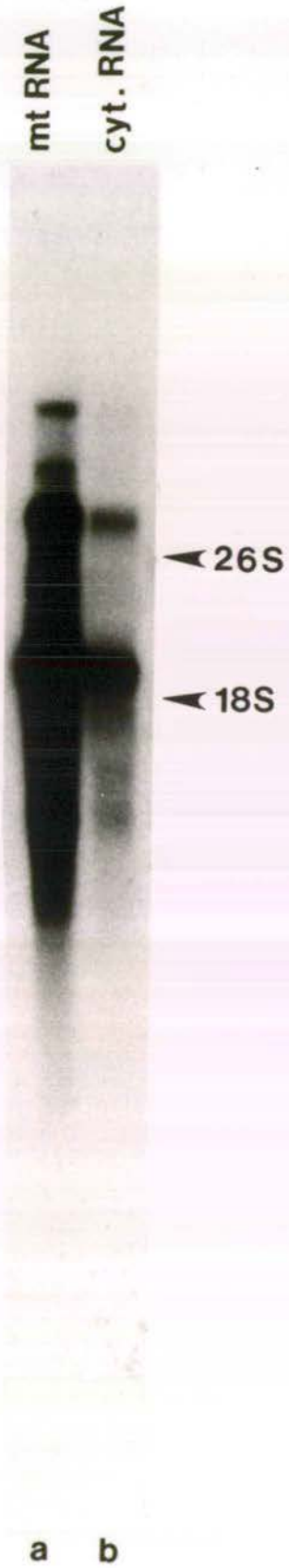


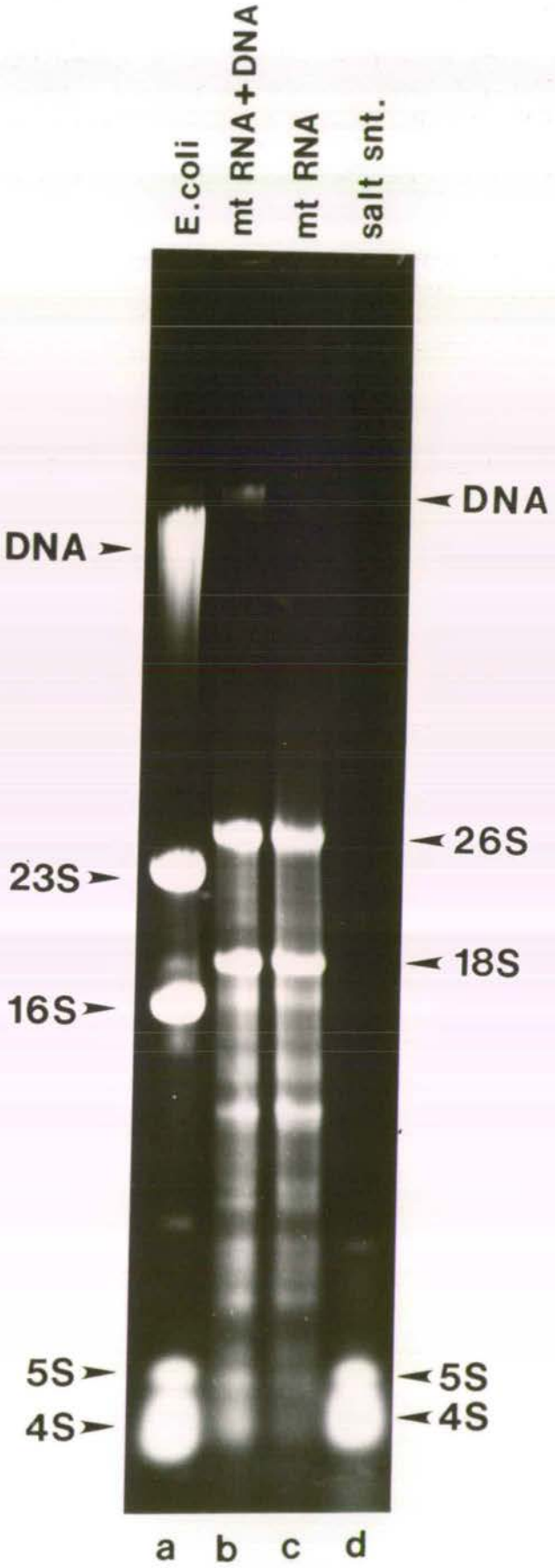
Fig. 4.5 Removal of DNA from preparations of maize mitochondrial nucleic acid

Total mt nucleic acids were treated with DNase, and RNA selectively precipitated by the addition of NaCl to a final concentration of 2.5M. Nucleic acids before and after DNase treatment were analysed by electrophoresis through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel and visualised by staining with ethidium bromide and visualised under U.V. light.

Track a) *E. coli* total nucleic acids (10µg)
Track b) Maize mt nucleic acids (10µg) before treatment with DNase
Track c) Maize mt RNA (10µg) after treatment with DNase
Track d) Analysis of the supernatant after salt precipitation

Note the absence of DNA (as judged by ethidium bromide fluorescence in track c. Also note no apparent alteration in the quality of RNA in track c (after DNase treatment) compared to track b (before DNase treatment).

Fig. 4.5


E.coli
mt RNA+DNA
mt RNA
salt snt.
DNA
DNA
23S
16S
26S
18S
5S
4S
5S
4S
a
b
c
d

DNA mono- and oligonucleotides generated by DNase digestion (Maniatis et al 1982) were removed by salt precipitation (Parish and Kirby 1966) which also removed 4S and 5S RNA. Treatment with DNase did not cause any detectable degradation of the RNA.

4.2.2.2 Plastid and cytoplasmic RNA contamination

Contamination of mt RNA with plastid and total cytoplasmic RNA was found to be insignificant because:

1) As discussed in section 4.2.1.1.2, E. coli DNA hybridises strongly to only two bands which comigrate with mt 26S and 18S rRNAs. Significant plastid RNA contamination would have been detected by hybridisation of chloroplast 23S and 16S rRNAs to the E. coli rDNA, since these share high homology (71% homology between maize chloroplast 23S and E. coli 23S; 74% homology between maize 16S and E. coli 16S. Schwarz and Kössel 1980, Edwards and Kössel 1981).

2) It is likely that there is some cytoplasmic RNA contamination since an RNA band which co-migrates with cytoplasmic 25S can usually be detected in electrophoretograms of maize mt RNA (not shown). However as shown in section 4.2.1.1.3, the presence of cytoplasmic RNA apparently does not influence the pattern of transcripts hybridising to the mt gene probes examined.

4.2.3 Gel electrophoresis of RNA and Northern blotting

As outlined in section 4.2, the technique of Northern blotting was chosen for the initial characterisation of maize mt transcripts.

4.2.3.1 Assessment of gel systems

In order to characterise accurately transcripts on the basis of their molecular weight, RNA molecules must be completely denatured, since:

1) mobility of native RNA in gels is related to both its molecular weight and its secondary and tertiary structure (Loening 1967) and

2) RNAs have a tendency to aggregate (Boedtker 1968, Grierson 1982).

The following three gel systems were assessed:

1) Agarose gels plus denaturing sample buffer

RNA was denatured by heating at 60°C for five minutes in a sample buffer containing 42% urea, and snap cooled on ice. The body of the gel contained no denaturant. Although these gels were useful for visualising RNA (because ethidium bromide could bind to the secondary structure in the RNA), they were not used for accurate molecular weight determinations, because the denaturation was temporary and renaturation and aggregation could occur on cooling or entering the gel in the absence of a denaturing agent actually in the gel.

2) Denaturation of RNA with glyoxal

Glyoxal (ethanedial) denatures RNA by covalently combining with guanosine residues and thus sterically hindering the formation of G - C base pairs (Shapiro and Hachman 1966). RNA was denatured by heating in 1M glyoxal and 50% (v/v) Me_2SO for 1 hour at 50°C (McMaster and Carmichael 1977). Since the covalent interaction is reversible only at high pH, the body of the gel did not contain denaturant. The use of this gel system allowed fairly accurate molecular weight determinations to be made. However, the long denaturation process (1 hour) was inconvenient and could lead to breakdown of RNA.

3) Agarose/formaldehyde gels

Formaldehyde agarose gel electrophoresis (Lehrach et al 1977) was the system eventually chosen for the analysis of maize mt RNAs. RNA was denatured by heating at 60°C for 5 minutes in 4% (v/v) formaldehyde/19% (v/v) formamide. Both formamide and formaldehyde denature

RNA by disrupting the hydrogen bonds (Grossmann 1968, Petermann et al 1972, Lodish 1971) that are involved in Watson-Crick base pairing. Renaturation is prevented by including formaldehyde in the body of the gel. A modification (Maniatis et al 1982) of the formaldehyde/ agarose gel system (Lehrach et al 1977) whereby NaPi buffer was substituted by MOPS gave better buffering capacity and made buffer recirculation or running of submerged gels unnecessary. These gels gave good resolution: in the range ca 1 - 3 Kb, bands differing by 100 bases could be resolved on a 1.3% (w/v) agarose gel. The disadvantage of these gels is that the denaturant must be removed before ethidium bromidecan bind to the RNA; this process takes several hours, thus some diffusion of the RNA bands occurs. In addition, residual formaldehyde was found to fluoresce in U.V. light, giving rise to a high background.

4.2.3.2 Blotting medium

Of the available blotting media [for example diazophenylthioether (DPT) paper (Seed 1982), "Biodyne" (Pall, Europe Ltd.), diazobezyloxymethyl (DBM) paper (Alwine et al 1977) and nitrocellulose] only the latter two were assessed.

1) DBM paper Bonds between phosphate groups on RNA and diazonium groups of the paper covalently attach RNA to DBM paper (Alwine et al 1977). This has the advantage that the RNA is permanently fixed on the support membrane. However this covalent coupling can be a disadvantage in that the number of binding sites is limited and regions of the filter may become saturated, particularly with the rRNAs. This is particularly important in the case of plant mt mRNAs where many of the transcripts are similar in size to the rRNAs and

may thus escape detection. The use of DBM paper was also time consuming and expensive, and the diazotization reaction sometimes worked with variable efficiency.

2) Nitrocellulose

The way in which RNA binds to nitrocellulose is not understood in detail, however it appears that RNA must be denatured for efficient binding to occur (Thomas 1980, 1983). Nitrocellulose is convenient to use (no pre-treatment of gel or membrane is required) and gives better resolution of individual transcripts than DBM paper. One major disadvantage is that, unlike DBM paper, staining of RNA prior to transfer reduces the efficiency of transfer (Thomas 1980), and molecular weight markers must be run in parallel tracks and excised from the gel prior to staining.

Whilst the three gel systems and two blotting media generally gave very similar results, the system used for most of the work described in this thesis consisted of formaldehyde/agarose gel electrophoresis followed by blotting to nitrocellulose. This method was convenient, gave good resolution and blots substantially free of any spurious background hybridisation.

4.2.3.3 Determination of the molecular weight of RNAs fractionated on formaldehyde/agarose gels

The approximate molecular weights of transcripts separated on formaldehyde/agarose gels were estimated by comparison with RNAs of known molecular weight, namely:

	Size bases	Reference
Tobacco mosaic virus (TMV)	6395	Goelet *et al* 1982
Cowpea chlorotic mottle virus (CCMV)	ca 3200 2900 2300 825	Davies and Verduin 1979 Dasgupta and Kaesberg 1982
E. coli 23S rRNA	2904	Brosuis *et al* 1980
16S rRNA	1541	Brosuis *et al* 1978
Maize mt 26S rRNA	ca 3546*	Dale *et al* 1984
18S rRNA	1955**	Spencer *et al* 1984

* From DNA sequence: actual length not known as 26S rRNA has not been sequenced

** Actual figure shown is for *wheat* rDNA

The lengths of the RNA markers have been determined directly by sequence analysis either of rRNA or rDNA, except for the three largest RNAs of CCMV which have been determined indirectly.

Figure 4.6 shows that the distance migrated is directly proportional to $\log_{10}$ molecular weight, except for very large RNAs, for example TMV = >6 Kb. Due to a lack of more than one marker greater than 3.5 Kb estimates of molecular weights greater than this must be regarded as only approximate.

4.3 TRANSCRIPT ANALYSIS OF THE MAIZE *COII* GENE

4.3.1 The *COII* gene

The protein coding portion of the maize *COII* gene is 825 bp in length and is interrupted by a single centrally located intron of 794 bp. The maize mt DNA contains a single complete copy of this gene (Fox and Leaver 1981). A preliminary investigation of the transcription of this gene was carried out by Fox and Leaver (1981). Using DNA

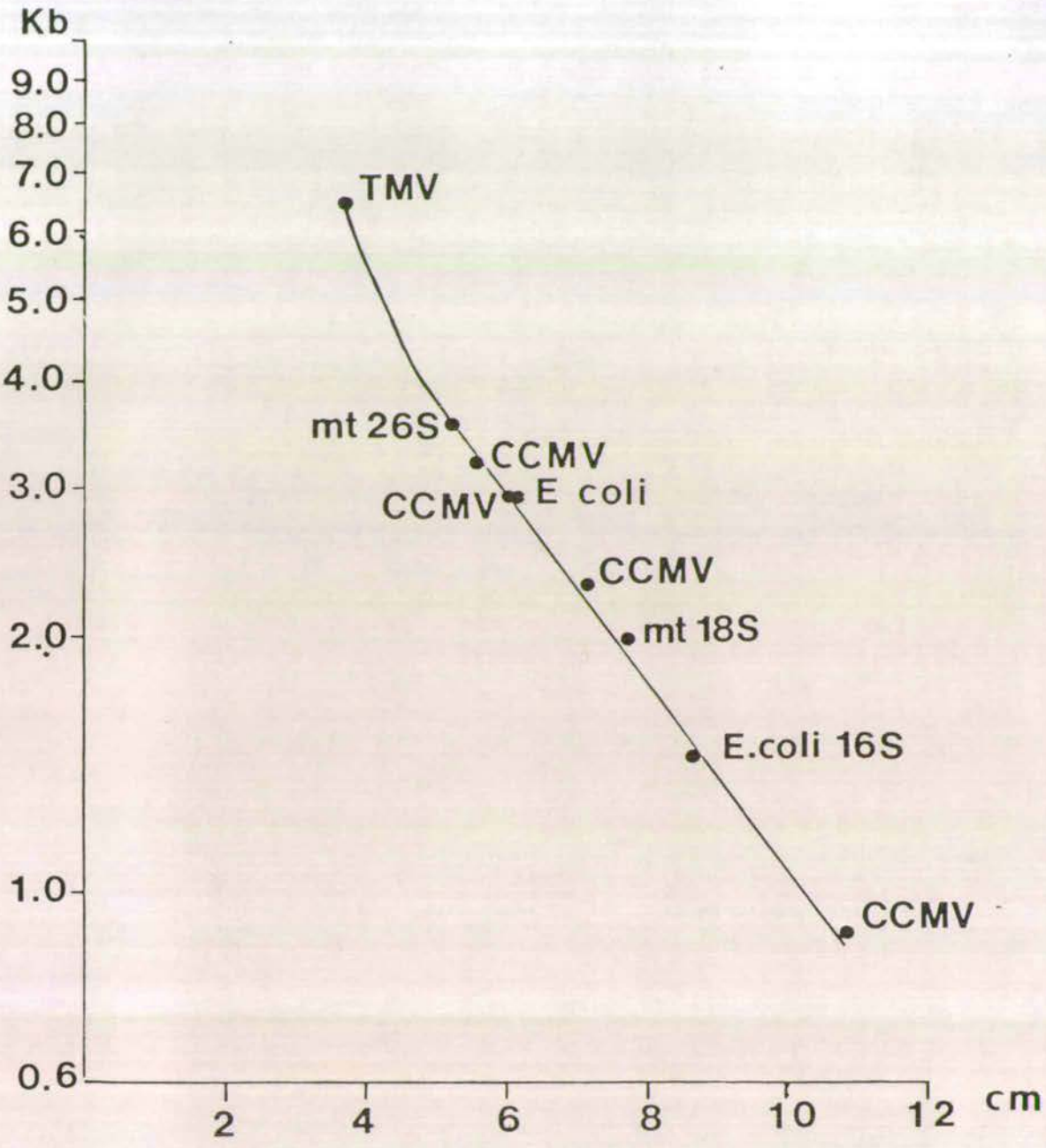


Fig 4.6 Plot of distance migrated against $\log_{10}$ molecular weight of RNA.

RNA was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel. The molecular weights of the RNAs were plotted against distance migrated in cms on a semi-logarithmic basis. Markers were: TMV (6395 bases), CCMV (ca. 3200, 2900, 2300 and 824 bases), E.coli rRNAs (2904 and 1541 bases), maize mt rRNAs (ca. 3546 and 1955 bases).

probes specific for exons 1, 2 and the intron, they showed that COII is apparently transcribed in a complex manner. However the probes used were not cloned fragments of DNA; rather they were extracted from a gel, and thus were liable to contamination with other DNA fragments.

In the work described in this section, single stranded M13 DNA probes were prepared and used to study the transcription of COII. In this way only transcripts generated from the COII coding strand were investigated. In addition,M13 probes can be labelled to high specific activity (5 - 10 x 10^6 dpm/μg) so allowing less abundant transcripts to be detected.

4.3.2 Construction of a single stranded COII intron-specific probe in M13mp8

The plasmid pZmE1 (Fox and Leaver 1981) contains a 2.4 Kb EcoRI fragment of mt DNA containing the whole of the COII gene cloned in pBR322. Plasmid DNA was digested with EcoRI and the insert separated from vector by electrophoresis through a 1% (w/v) agarose gel. The insert was recovered from the gel by electroelution and 4 μg digested with TaqI. TaqI was chosen because it generates a large intron-specific fragment of 692 bp which is well-separated from other fragments (Fig. 4.7). The products of the TaqI digestion were separated on a 6% (w/v) polyacrylamide gel (Fig. 4.8). The desired fragment was recovered from the gel by 'crush and soak' and ligated into the AccI site of M13mp8. The ligated DNA was used to transform competent E. coli JM101, and RF and ss DNA prepared from a random selection of transformants. Recombinant transformants were screened initially by measuring insert size. An aliquot of RF DNA was

Fig. 4.7 Restriction map of pZmE1 insert

Restriction map of the 2.4 Kb insert in pZmE1 showing TaqI sites (T) and EcoRI sites (E). Arrows delimit the 692 bp TaqI fragment cloned into M13mp8 to form an intron specific probe.

Thickened bar denotes gene coding sequence.
Solid thickened bars represent exons, open thickened bar represents the intron.

Fig. 4.8 Purification of the 692 bp TaqI intron specific fragment

pZmE1 insert (4µg) was digested with TaqI and electrophoresed on a 6% (w/v) polyacrylamide gel. Bands were visualised under incident U.V. light after staining with ethidium bromide. pBR322 digested with TaqI provided markers.

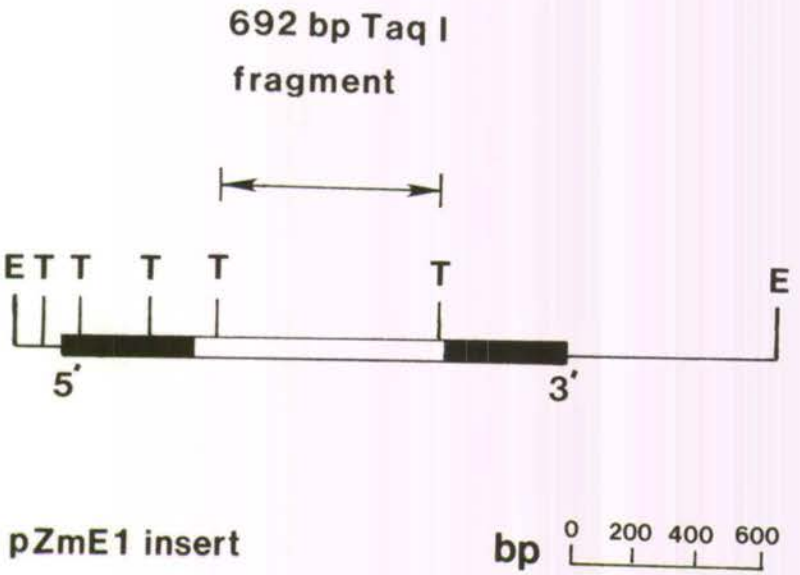

692 bp Taq I
fragment
E T T T T T E
5'
3'
pZmE1 insert
bp 0 200 400 600

Fig.4.7

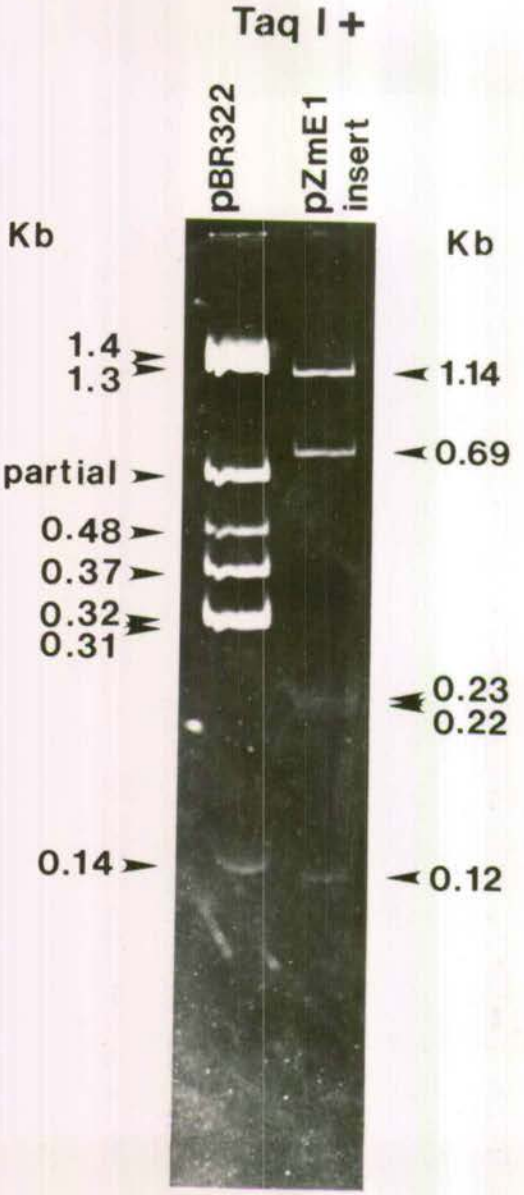

Taq I +
pBR322
pZmE1
insert
Kb
Kb
1.4
1.3
partial
0.48
0.37
0.32
0.31
0.14
1.14
0.69
0.23
0.22
0.12

Fig.4.8

was digested with HindIII and EcoRI which cut either side of the AccI site in the polyclonal linker. Clones containing the 692 bp TaqI intron fragment were thus contained HindIII - EcoRI inserts of 728 bp. Figure 4.9 shows clones 4, 5, 7 and 8 contain correctly sized inserts. Nucleotide sequencing established the orientation of these clones. The products of a sequencing reaction of clone 4 are shown in Figure 4.10. The orientation of clones 4, 5 and 8 is such that the labelled strand is complementary to the RNA and will therefore hybridise to any RNA transcribed from this region.

4.3.3 Construction of a single stranded COII exon 1 probe in M13mp8

A single stranded M13 DNA probe for exon 1 of COII was prepared as follows: pZmE1 was digested with Sau3A (a restriction map of the mt DNA insert in pZmE1 is shown in Fig. 4.11) and the products electrophoresed on a 6% (w/v) polyacrylamide gel (Fig. 4.12). Bands around 269 bp were excised and ligated into the BamHI site of M13mp8. After transformation of E. coli JM101, a random selection of recombinants were sequenced directly, the results of which are shown in Fig. 4.13. Two clones (47 and 50) were found to be correctly oriented for hybridisation with RNA.

4.3.4 Transcript analysis of the maize COII gene using internal gene DNA probes

Mt RNA was prepared from etiolated seedlings and 50µg loaded into a single well 5cm wide. Following electrophoresis through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel the RNA was immediately transferred to nitrocellulose. The RNA which was blotted was not stained prior to transfer; RNA size markers were run in parallel tracks and stained separately.

Fig. 4.9 Screening recombinant M13 clones by measurement of insert size

Recombinant transformants were screened by measuring insert size. Clones containing the 692bp TaqI intron fragment were identified by restriction with HindIII and EcoRI which cut either side of the AccI site in the mp8 polyclonal linker. Clones containing the 692bp TaqI intron fragment thus contained HindIII - EcoRI inserts of 728bp. RF DNA was digested with EcoRI and HindIII and electrophoresed on a 1% (w/v) agarose gel. Bands were visualised under U.V. light after staining with ethidium bromide. pBR322 and TaqI provided markers (M).

Fig. 4.10 Nucleotide sequence of clone 4

Sequencing gel of clone 4
Clones correctly oriented for hybridisation with RNA should read:
5' CGCCCGCCTCCCCCCAAA 3'

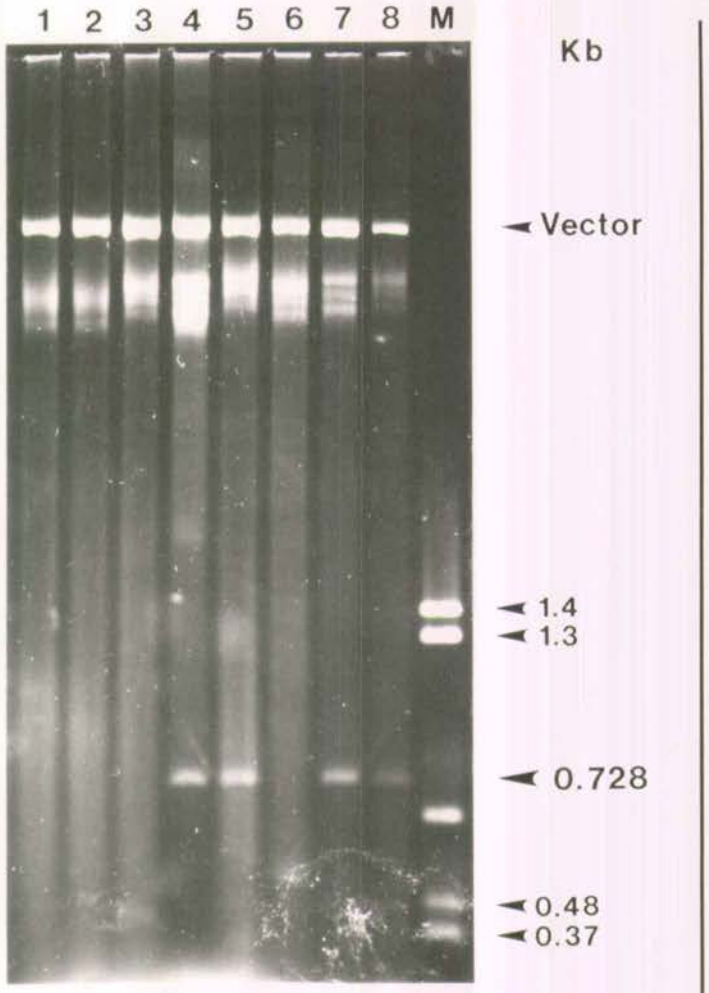


Fig. 4.9

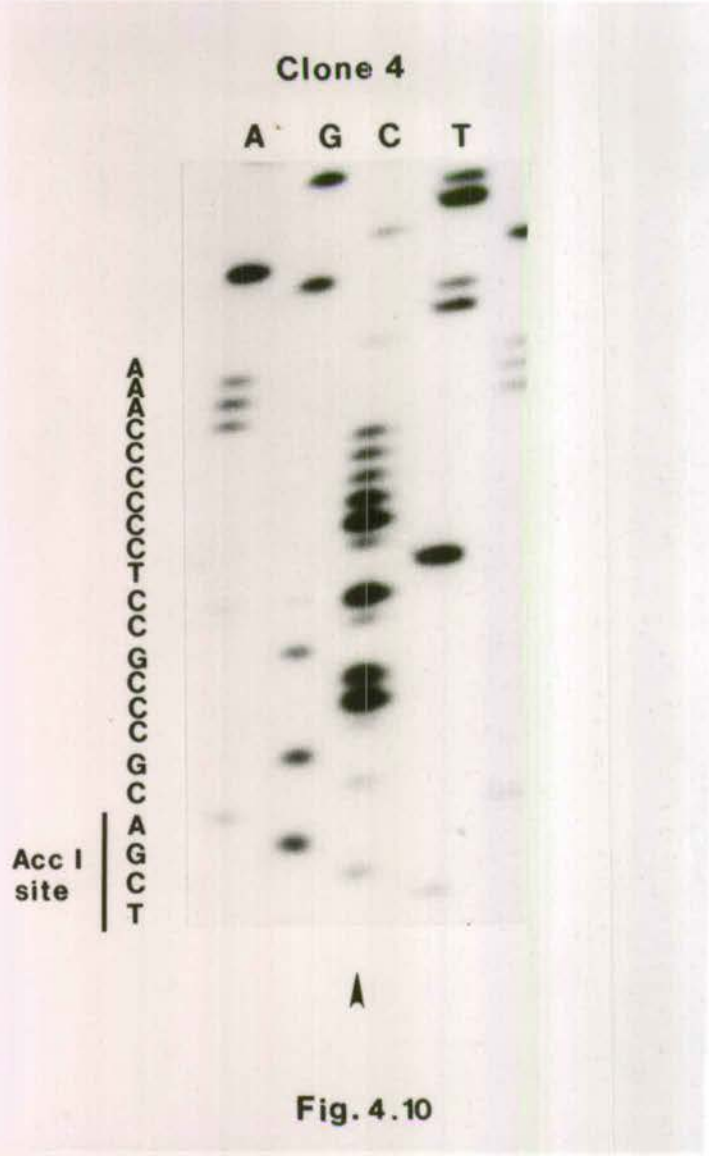


Fig. 4.10

Fig. 4.11 Restriction map of pZmE1 insert

Restriction map of pZmE1 insert to show Sau3A restriction enzyme sites (S) and EcoRI sites (E). Thickened bars represent gene sequence. Solid thickened bars represent the two exons. Open thickened bar represents the central 794 bp intron. Arrows delimit the 269 bp Sau3A exon 1 specific probe.

Fig. 4.12 Purification of the 269 bp Sau3A exon 1 specific fragment

pZmE1 (4µg) was digested with Sau3A and the fragments fractionated by electrophoresis through a 6% (w/v) polyacrylamide gel. DNA bands were visualised under incident U.V. light after staining with ethidium bromide. pBR322 digested with Sau3A provided markers.

Fig. 4.13 Nucleotide sequence of clone 47 (COII exon 1 specific clone)

Sequencing gel of clone 47. Clones correctly oriented for hybridisation with RNA should read 5' TACTAATACCCCGTCCATTGAG 3'

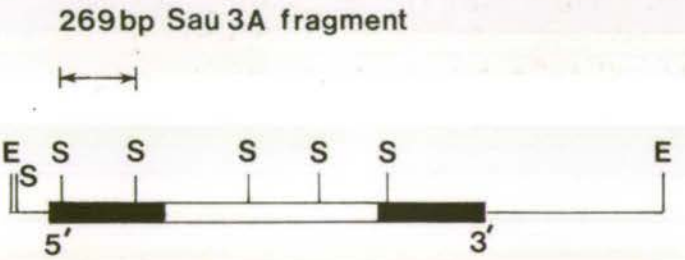


Fig. 4.11

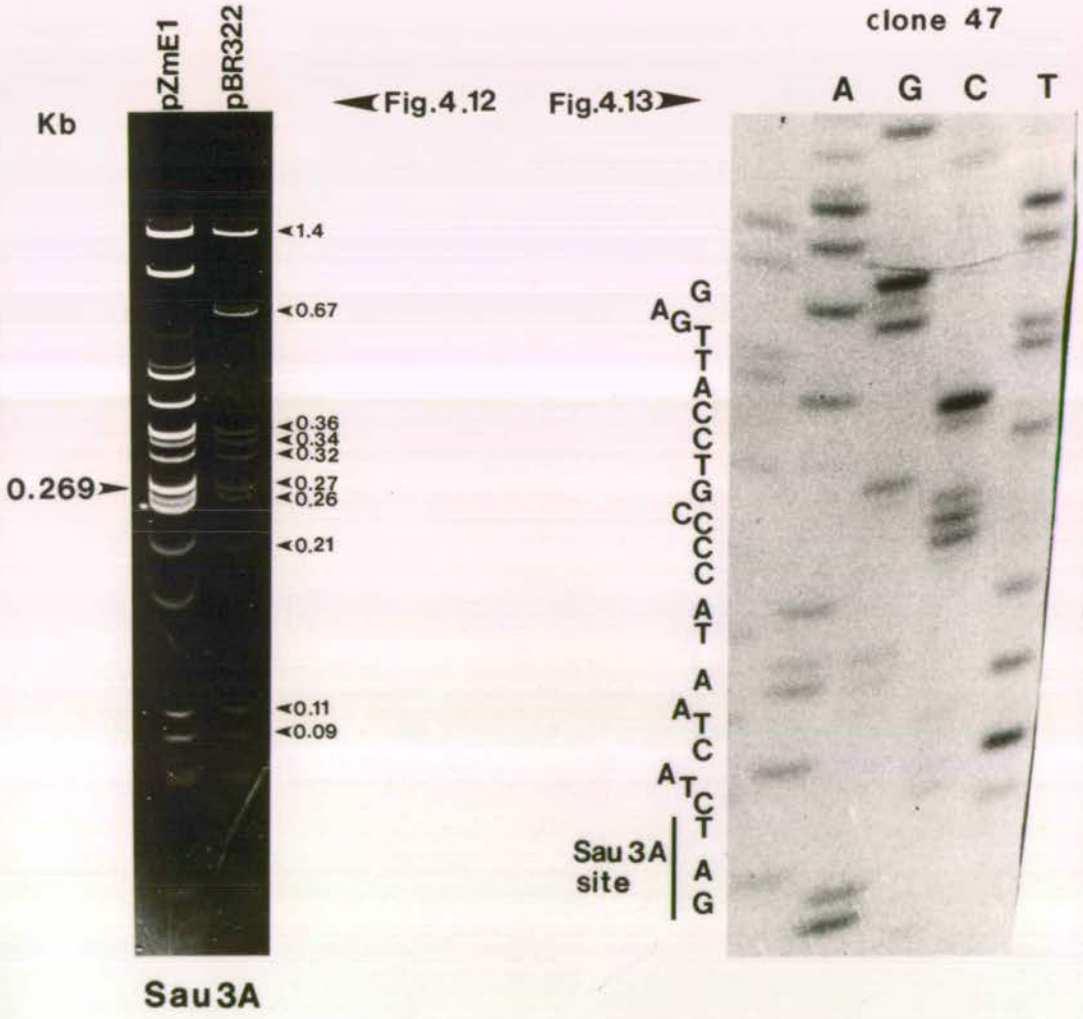


Fig.4.12

Fig.4.13

Figure 4.14 a shows these RNA markers stained with ethidium bromide and visualised under incident U.V. light. (Note the apparent loss of 26S rRNA from the mt RNA track is due to uneven illumination.) Parallel strips from the nitrocellulose blot were excised and hybridised with the following probes:

1) Nick-translated pZmE1, a double stranded DNA probe containing the entire COII gene and approximately 800bp of flanking sequence (100bp 5' of the gene and 700bp 3').

2) The M13 COII exon 1 specific probe (section 4.3.3) labelled to 5 - 10 x 10^6 dpm/μg by second strand synthesis.

3) The M13 COII intron specific probe (section 4.3.2) labelled to high specific activity as in 2) above.

The results of these hybridisation experiments are shown in Fig. 4.14 b c and d. The use of identical blots of mt RNA allows direct comparison to be made between the RNA transcripts hybridising to each probe.

The pattern of transcripts hybridising to pZmE1 and exon 1 specific probe are identical and complex (lanes b and c Fig. 4.14).
This shows that all the transcripts identified are complementary to the COII coding strand. It also indicates that neither M13 nor pBR322 vector homology makes any significant contribution to the transcript pattern.

The pattern of transcripts hybridising is unexpectedly complex. At least five transcripts ranging in size from ca 1.0 to ca 3.2 Kb hybridise strongly to the probe. The most abundant transcripts of 1.95 and 3.2 Kb are presumed to be the mature mRNAs, but this may not be a valid assumption since one or both may represent a stable

Fig 4.14 Analysis of maize COII gene transcripts

Maize mt RNA was electrophoresed through a 1.3% (w/v) agarose/ 7% (v/v) formaldehyde gel, blotted to nitrocellulose, and hybridised with DNA probes from specific regions of the COII gene.

4.14 a RNA markers were run in parallel tracks, excised from the gel after electrophoresis and stained with ethidium bromide separately. RNA markers are cowpea chlorotic mottle virus (CCMV), E. coli, maize mt RNA and tobacco mosaic virus (TMV).

4.14 b Transcripts hybridising to pZmE1 (entire nick translated plasmid). Open arrows denote weakly hybridising bands not reproduced from the original.

4.14 c Transcripts hybridising to the COII exon 1 specific probe. Open arrows denote weakly hybridising bands.

4.14 d Transcripts hybridising to the COII intron specific DNA probe.

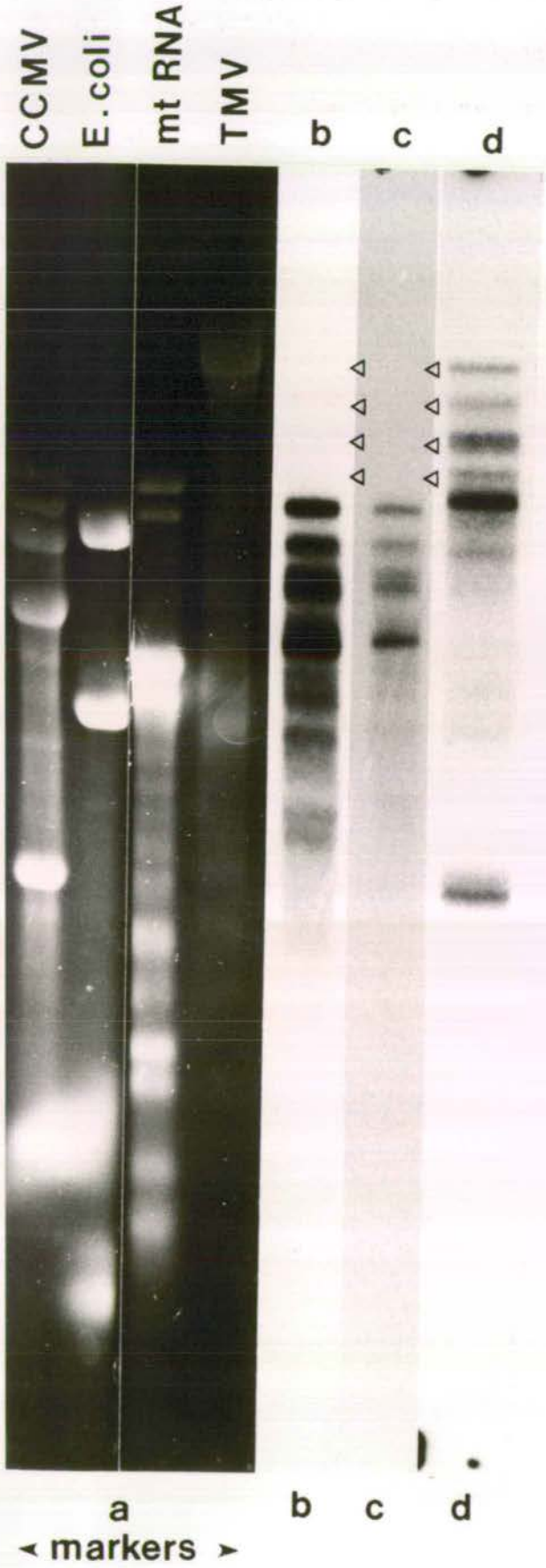


Fig. 4.14

processing intermediate. In addition to the strongly hybridising bands, a number of weakly hybridising bands of higher molecular weight can be detected. (Denoted by open arrows on Fig. 4.14 b and c.) The intron-specific probe hybridises strongly to at least five high molecular weight transcripts ranging in size from ca 3.5 Kb to 6.0 Kb (Fig. 4.14 d). These co-align with the high molecular weight transcripts that hybridise faintly to pZmE1 and exon 1-specific probe. These presumably therefore are unspliced pre-mRNAs. In addition to high molecular weight transcripts, the intron-specific probe hybridises strongly to a low molecular weight RNA species of approximately 800 bases, approximately corresponding in size to the 794bp intron. Also a number of faintly hybridising transcripts of less than 3.5 Kb but greater than 800 bases are detected with the intron-specific probe. The significance of these is not known.

4.3.5 COII transcripts in N C S and T type mitochondria

There is evidence that the COII gene undergoes rearrangements with respect to N in C and T type mitochondria. These are discussed in section 4.3.6.6. Maize mt RNA prepared from N (nuclear genotype 3541), and N, C, S and T (nuclear genotype WF9) type plants was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, blotted to nitrocellulose and hybridised with the COII exon 1 specific probe (Fig. 4.15). COII transcripts from N cytoplasm in the nuclear backgrounds 3541 and WF9 are identical; hence the nuclear background does not contribute to any change in COII transcripts. Differences in the pattern of transcripts are detected in all three male sterile (C, S and T) cytoplasms, the most obvious changes being in C and S type mt RNA.

Fig. 4.15 COII transcripts in N C S and T mt RNA

Mt RNA from a variety of cytoplasmic genotypes and two nuclear backgrounds was electrophoresed through a 1.3% (w/v) agarose/ 7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with the COII M13 exon 1 specific clone.

Track a) N cytoplasm, nuclear background 3541
Track b) N cytoplasm, nuclear background WF9
Track c) C cytoplasm, nuclear background WF9
Track d) S cytoplasm, nuclear background WF9
Track e) T cytoplasm, nuclear background WF9

The positions of mt 26S and 18S rRNAs are marked. These were determined by staining of the nitrocellulose with ethidium bromide.

Fig. 4.15

4.3.6 Discussion

4.3.6.1 Summary of results

Transcript analysis of the maize COII gene has shown:

1) Mt RNA hybridises to COII specific probes, indicating that the gene is probably transcribed in the maize mitochondrion.

2) A complex pattern of transcripts hybridises to COII specific probes ranging in molecular weight from ca 800 bases - 6 Kb. The large number of transcripts cannot be explained by excision of a single intron. The majority of COII transcripts are far larger than required to specify the corresponding polypeptide (825bp).

3) The intron specific probe hybridises to a low molecular weight RNA of ca 800 bases suggesting it may be spliced from the pre-mRNA(s) and persist in the mitochondrion as a stable molecule.

4) The pattern of COII transcripts differs in C S and T type mt RNA compared to N type mt RNA.

4.3.6.2 Multiple COII transcripts

Several models can be proposed to explain the presence of multiple transcripts which hybridise to COII specific probes:

1. Initiation and/or termination of transcription occurs at many sites throughout the COII region thus generating a series of overlapping transcripts. In this case the relative abundance of individual transcripts would depend on the efficiency of the competing promoters and terminators, as well as the rate of turnover of the individual transcripts. Two putative promoters have been found 630bp and 552bp upstream of the yeast mitochondrial ATPase 9 gene (Edwards et al 1983); similarly the rRNA genes in E. coli (Lund and Dahlberg 1979) and the E. coli gal operon (Queen and Rosenberg

1981) are expressed from tandem promoters.

2. A single precursor transcript (for Zea mays COII ca 6 Kb) is processed at many sites to yield transcripts of lower molecular weight, some of which are processing intermediates, others mature mRNAs. Here the relative abundance of individual transcripts would depend on the rate of each processing step and the stability of the resultant processed intermediate. Thus the Zea mays COII primary transcript would undergo two types of processing; excision of the 794bp intron, and trimming of the 5' and 3' ends to form the mature mRNA. If this is so, then the detection of several high molecular weight transcripts containing intron sequences suggests some trimming may precede intron excision (although the intron could be excised in stages). It is also possible that the same precursor RNAs are processed by different pathways. The complex pattern of Zea mays COII transcripts resembles that of some yeast genes which contain introns (eg COI, Hensgens et al 1983). On the other hand,DNA probes for the yeast mt COII gene (which is not split) hybridise to at least 10 discrete yeast mt RNA species, ranging in size from ca 4.0 Kb to the presumed mature mRNA of 0.85 Kb. The pattern of hybridisation obtained using probes flanking the protein coding sequence suggests that a larger precursor is processed in a stepwise fashion from both 5' and 3' ends to generate the mature mRNA (Co ruzzi et al 1981). Thus a single primary precursor maize COII transcript could be processed by such a multi-step pathway.

3. A combination of 1 and 2.

From the data presented in this section it is impossible to distinguish between these three models, particularly since the positions of the COII transcripts with respect to the protein coding sequence have

not been determined. Several approaches could be used to distinguish the three models:

1) Initiation of transcription at multiple points upstream from the protein coding sequence requires that there must be multiple promoters which presumably share sequence homology, as do yeast putative promoters (Tabak et al 1983). Using the S1 nuclease protection technique (Berk and Sharp 1977) the sequences around the 5' and 3' ends of individual transcripts could be determined and examined for sequence similarities. The multiple promoter/terminator model predicts conserved sequences immediately 5' to the point at which transcription is initiated and/or at the 3' termini of the transcripts. A more ambitious approach would be to examine the products of in vitro transcription using various DNA templates and a purified mt RNA polymerase. Such a system is used routinely in the study of yeast mt transcription (Edwards et al 1982) but an equivalent system has not yet been developed for plant mt DNA.

2) In contrast, processing of a single primary transcript by a complex pathway does not necessarily predict any sequence homology at the transcript termini since the different transcripts could be generated by various processing enzymes. Alternatively processing signals may be situated distantly from the actual site of cleavage.

3) The fact that the plant mt genome undergoes recombination at various repeated sequences (Palmer and Shields 1984, Lonsdale et al 1983) has important implications for the study of mt gene expression. If, for example, a gene is located near to a repeated element which is involved in recombination upstream/downstream sequences could be altered completely thus having a profound effect on transcription. The Zea mays COII gene lies within a few hundred bases of a 2 Kb

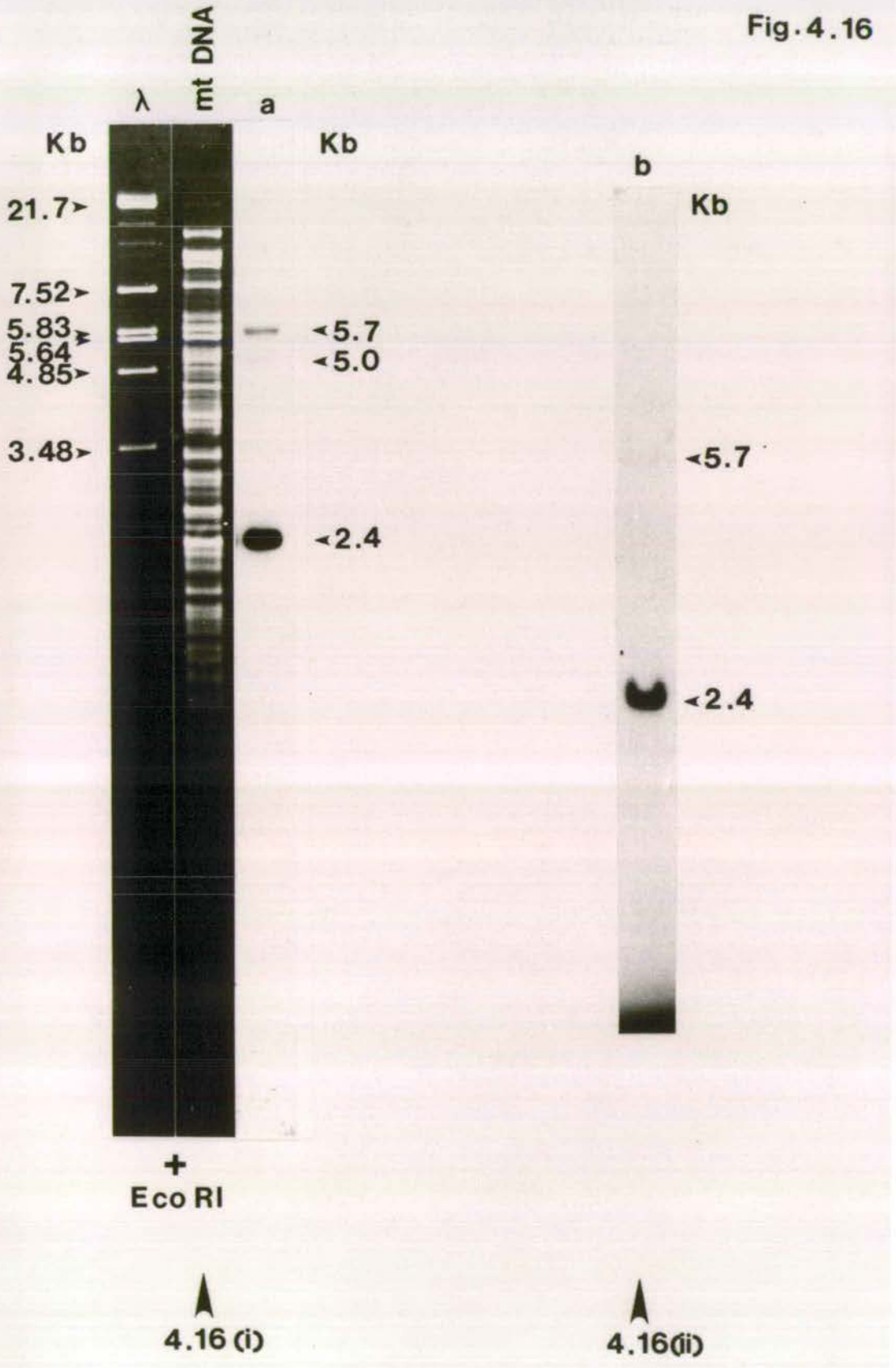
repeated element which has been shown to be involved in recombination events (D. Lonsdale pers. commun.). However, since the start point of the primary COII transcript is not known, it is impossible to make any predictions about the effect of recombination on transcription. It is possible that the multiple COII transcripts could arise from repeated COII sequences. This was originally thought to be unlikely since nick-translated pZmE1 hybridises to a single 2.4 Kb EcoRI fragment when used as a probe against N-type mt DNA (Fox and Leaver 1981). In order to verify that only a single copy of COII exists, N type mt DNA (prepared as described in Dawson et al 1984) was digested with EcoRI, electrophoresed through a 1% (w/v) agarose gel, blotted to nitrocellulose (blot kindly provided by Sue Dunbar) and hybridised with the COII exon 1 specific M13 clone probe (section 4.3.3) which was labelled to high specific activity (5 - 10 x 10^6 dpm per µg). Since this probe contains only COII specific sequence it should offer greater sensitivity in detecting COII sequences compared to pZmE1 which contains 800bp of flanking sequence. Fig. 4.16 (i) shows the results of this hybridisation. Under stringent conditions (T_m DNA - 26°C) the exon specific probe hybridises to a 2.4 Kb fragment and faintly to several other fragments of 5.7 and 5.0 Kb. Re-examination of a different blot of EcoRI digested maize mt DNA probed with pZmE1 showed that a band of ca 5.7 Kb hybridises to the probe [autoradiogram provided by A. Dawson; Fig. 4.16 (ii)]. Thus the weakly hybridising bands are only readily detected using highly specific probes. Under these conditions M13mp8 vector does not hybridise to mt DNA (data not shown). The additional bands are not likely to be explained by partial digestion products (since the same hybridisation has been detected with several different restriction digests of mt DNA) or methylation of

Fig. 4.16 Repeated COII sequences in the mt genome of maize

4.16 (i) Maize mt DNA was electrophoresed through a 1% (w/v) agarose gel, transferred to nitrocellulose and hybridised with the COII exon 1 specific clone in M13. Lane a shows results of this hybridisation experiment aligned with the corresponding mt DNA visualised with ethidium bromide. λ DNA digested with EcoRI provided markers. Note weakly hybridising bands of 5.7 and 5.0 Kb.

4.16 (ii) Maize mt DNA (separate preparation from 4.16 (i)) was digested with EcoRI, electrophoresed through a 1% (w/v) agarose gel, blotted to nitrocellulose, and hybridised with nick translated pZmE1.
Note faint band at 5.7 Kb hybridising.

Fig. 4.16


λ
mt DNA
a
Kb
Kb
21.7
7.52
5.83
5.64
4.85
3.48
5.7
5.0
2.4
+
Eco RI
4.16 (i)
b
Kb
5.7
2.4
4.16(ii)

of restriction sites (Bonen *et al* 1980). Thus it appears that *COII* homologous sequences are repeated on the maize mt genome. In view of the weak hybridisation compared to the 2.4 Kb fragment, the nature of the repeat could be:

1) Short pieces of sequence completely homologous to *COII* exon 1

2) Longer sequences having partial homology to *COII* exon 1

3) Complete copies of *COII* on submolar restriction fragments which could arise from rare recombination events.

Sequences showing only partial homology to *COII* could be identified by increasing stringency of the wash, but this has not yet been investigated. Further experiments using intron and exon 2 specific probes are necessary to determine the extent of the *COII* gene that is repeated on the maize mt genome. Bonen *et al* (1984) have since shown that in wheat, 193bp of *COII* exon 1 are repeated elsewhere in the wheat mt genome and thus repeated gene sequences may be a feature of plant mt mt DNAs.

If the maize repeated *COII* sequence(s) are transcribed (whether they form part of a functional *COII* mRNA or part of some other transcript), then they could contribute to the overall complexity of the *COII* transcription pattern. In order to determine whether these repeated *COII* sequences are transcribed, it would be necessary to:

1) hybridise DNA probes flanking the additional copies (or copy) of *COII* to blots of mt RNA;

2) examine the *in vitro* transcription of the repeated *COII* sequences, or

3) identify primary *COII* transcripts using the enzyme guanylyl transferase (section 4.4.6.3), sequence the resultant 'capped' RNAs

and align the sequence to the genome. These experiments are described in more detail in section 4.4.6.3.

Recently, transcripts hybridising to the wheat mt COII gene have been identified (Bonen et al 1984). This gene contains a centrally located intron of 1.2 Kb; the majority of the intron shows very high homology ($>$ 99%) to the maize 794bp COII intron. (The extra sequence is due to a 422bp insert within the intron). The pattern of RNAs hybridising to this gene are very simple: three transcripts of 2.7, 1.5 and 1.2 Kb are detected corresponding to the unspliced pre-mRNA, mRNA, and the stable intron species respectively. Thus a complex pattern of transcription as seen for maize COII gene does not necessarily result from the presence of an intron. Also, the finding that 193bp of COII exon 1 is repeated once elswhere on the wheat mt genome shows that at least in this species, repeated COII sequences do not significantly contribute to the transcript pattern.

4.3.6.3 Processing of the COII intron from precursor mRNA

The identification of high molecular weight transcripts containing intron and exon sequences, together with the detection of a low molecular weight molecule of ca 800 bases which hybridises to the intron specific probe suggests that the intron may be spliced from the pre mRNAs and persist in the mitochondrion as a stable molecule. It has been proposed that introns can be spliced from precursor mRNAs by one of several mechanisms (for reviews see Cech 1983, Greer and Abelson 1984):

1. Nuclear encoded mRNAs whose intron borders show the consensus sequence 5'GT ↑ .... AG ↑ 3'(splice points are designated by arrows) and

whose splicing depends on the action of other cofactors, for example small nuclear ribonucleo-protein particles (Padgett et al 1983, Breathnach et al 1978). In this group, deletion of most of the intron does not affect splicing (Cech 1983).

2. Nuclear encoded tRNAs where introns interrupt the tRNA sequence in the anticodon loop one base 3' to the anticodon. In yeast a single endonuclease is involved in splicing of introns from ten tRNAs (Peebles et al 1983).

3. Mitochondrial and nuclear encoded rRNAs of a few species, and mitochondrial mRNAs. These are discussed in more detail below.

4. Introns in chloroplast genes. The genes encoding the large subunit of RuBPCase and the M_r 32,000 polypeptide in Euglena gracilis are split by nine and four introns respectively (Koller et al 1984, Karabin et al 1984). Sequence analysis of these introns has revealed a conserved sequence at the splice sites which is

$$5'\ NGTG^{T}_{C}G\ \overset{\text{intron}}{\ldots\ldots\ldots}\ TTTA^{A}_{G}TTTTATN\ .$$

Excision of mitochondrial introns in fungi is dependent on many factors:

1) proteins encoded in the nucleus (see for example Dujardin et al 1983, Garriga et al 1984, Faye and Simon 1983, Simon and Faye 1984)

2) intron encoded proteins (maturases) (see for example Lazowska et al 1980, Dujardin et al 1982, Anziano et al 1983)

3) specific sequences within the introns themselves which are responsible for generating the splice sites and complex internal secondary structure. (Davies et al 1982, Waring and Davies 1984.) Mitochondrial introns are divided into Groups I or II on the basis of this predicted secondary structure (Michel et al 1982, Michel and Dujon 1983).

Group I introns contain four conserved sequences capable of forming secondary structures which bring the 5' and 3' ends of the intron together and allow an 'internal guide RNA' sequence to pair with the exon bases adjacent to the splice junctions. This is postulated to bring the two splice sites into close proximity. Evidence for the formation of such a core structure comes from the finding that mutations in the conserved sequences are often compensated by secondary mutations so as not to disturb the secondary structure (Davies et al 1982). Examples of Group I introns include intron 4 of S. cerevisiae COB gene, (Rodel et al 1983), intron of large rRNA, and introns 1, 2 and 3 of A. nidulans (Davies et al 1983).

These Group I introns are also found in Tetrahymena and Physarum nuclear encoded large rRNA genes (Waring and Davies 1984, Cech et al 1983). Excision of the single intron from Tetrahymena 26S rRNA is an autocatalytic event, requiring no enzyme or protein in vitro (Kruger et al 1982). This intron splicing reaction excises a linear molecule which is subsequently circularised (Grabowski et al 1981).

RNA cleavage and ligation activity is therefore intrinsic to the structure of the RNA molecule. The close conservation of secondary structure of yeast mt Group I introns with that of Tetrahymena 26S rRNA suggests all these introns are or have been self splicing. The role of the maturases and nuclear encoded proteins involved in splicing of group I introns in yeast mt genes may be to stabilise secondary structure in some way.

Introns in Group II share extensive homologies at their 3' ends, and can form potential secondary structures different from those in Group I (Michel and Dujon 1983). Group II introns identified so far

include in S. cerevisiae intron 1 of COB, introns 1, 2 and 5 of COI and in Zea mays COII intron. The first four introns are excised as covalently closed circular molecules (Arnberg et al 1980, Halbreich et al 1980) and remain in the mitochondrion. It is not known whether the apparently stable spliced maize COII intron is excised as a circular molecule. As yet there is no experimental evidence for the existence of a circular spliced COII intron. However, there is some circumstantial evidence that it may exist in this form:

1) EM preparations of maize mt RNA do contain covalently closed circular RNA molecules of approximately 800 bases (A. Arnberg, University of Groningen and V.P. Jones unpublished results). By enriching for these circular RNA molecules (for example by sucrose gradient density centrifugation or elution from gels) and hybridising them with the intron specific probe, it should be possible to demonstrate whether these circular molecules arise from splicing of COII or not. Another approach would be by primer extension (Ghosh et al 1978). This involves hybridising a 5' ^{32}P-labelled primer (either a restriction fragment or a synthetic primer) to the intron and synthesising a cDNA strand using reverse transcriptase. Circularity would be indicated by a linking of 5' and 3' sequences. Using this approach the circular introns of S. cerevisiae have been shown to be closed by an unusual bond, since reverse transcriptase is unable to progress across the closure. The bond is not cleaved by pyrophosphatase, alkaline phosphatase or protease (L.A. Grivell pers. commun.). This suggests either that the circles may be closed and the bases subsequently modified in some way or that the RNA around the point of closure forms a branched structure.

2) The mobility of circular RNA species in denaturing RNA gels is

known to be anomalous. For example the circular RNA resulting from excision of the Tetrahymena 26S rRNA intron was detected as an electrophoretic variant from the linear form and was shown to be retarded on 8M urea polyacrylamide gradient gels (Grabowski et al 1981), increasing its apparent molecular weight by ca 100 bases. On non-denaturing acrylamide gradient gels it was found to migrate faster than the linear form, the latter having a molecular weight corresponding to that of the intron. Similar anomalous rates of migration have been found for circularised tRNAs (Bruce and Uhlenbeck 1978) and the circular RNA of Potato spindle tuber viroid (Sanger et al 1979).

The close agreement of the size of the excised maize COII intron and its DNA parent could mean either that the intron exists as a linear molecule, or that the resolving power of the gel system used is insufficient to detect minor differences in mobility due to circularisation. In wheat the COII intron excised RNA can be detected in Northern blots, although its physical conformation has not been investigated.

3) In yeast mitochondria only circularised excised intron RNAs only ie COB intron 1 and COI introns 1, 2 and 5 can be detected as apparently stable molecules in the mitochondrion (Bonitz et al 1980, Hensgens et al 1983). Other introns appear to be degraded more rapidly after excision from the pre mRNA and are not detected; thus the 800 base RNA species hybridising to Z. mays COII intron may represent a circular moiety.

The function of these circular RNAs is not known, although it appears that they may not be direct products of the splicing reaction since:

a) the Tetrahymena 26S rRNA intron is excised as a linear molecule which is subsequently circularised (Grabowski et al 1981) and

b) circle formation in yeast is thought to be uncoupled from splicing during glucose repression (Baldacci and Zennaro 1982) and sporulation (Schroeder et al 1983) since under these conditions circular transcripts are not detected. (Alternatively the activity of a circle specific RNase may increase under particular physiological conditions).

4.3.6.4 An open reading frame upstream of the maize COII gene

Sequence analysis of the COII gene upstream from the AUG initiation codon has shown the presence of an open reading frame (ORF) continuous with the COII gene. No termination codons separate this ORF from that encoding COII, and the ORF extends through the first exon and terminates 33 amino acids into the intron. The 5' start point of the upstream ORF has not yet been determined but extends at least 209 amino acid residues from the initation codon (A.J. Dawson pers. commun.). There are several possible roles this extension may have:

1) It may not be expressed, and although the 5' ends of COII transcripts have not been mapped, they are large enough to accommodate the entire COII coding sequence and that of the ORF.

2) The ORF may be co-translated with COII to form an amino-terminal extension which is subsequently processed from the mature polypeptide. In yeast (Sevarino and Poynton 1980) and Neurospora crassa (van den Boogaart et al 1982) COII polypeptides are synthesised with amino terminal extensionsof 12 - 15 amino acids. Similarly the COI genes in N. crassa (de Jonge and de Vries 1983) and A. nidulans (T. Brown pers. commun.) are thought to be translated as a precursor, with an amino terminal extension of around 50 amino acids.

3) The ORF may specify part of a maturase enzyme cotranslated with COII exon 1 and part of the intron, which is involved in splicing of

the intron from the pre-mRNA(s). As yet it is not known whether the maize COII gene intron is excised by a nuclear encoded enzyme or a mitochondrially encoded maturase.

In yeast, maturases translated in frame with the preceding intron include proteins partly encoded by introns 2, 3 and 4 of COB (Lazowska et al 1980) and introns 1 and 4 of COI (Groudinsky et al 1983). Thus both Group I and Group II introns partly encode maturases. Interestingly the ORF found upstream of the maize COII gene is not found in the homologous position in wheat mt DNA, (Bonen et al 1984) nor Oenothera mt DNA (Hiesel and Brennicke 1983). The Oenothera COII gene does not contain an intron, and therefore if the ORF was part of a maturase its absence would not be unexpected. The wheat COII gene is split by an intron, but shows a very simple pattern of transcription unlike the corresponding gene in maize. It is possible that the presence of the ORF in maize is contributing to the complex pattern COII RNAs revealed by hybridisation.

4.3.6.5 Large size of COII transcripts

Both pZmE1 and exon 1 specific probes hybridise to many RNA species, the majority of which are much larger than required to specify the COII polypeptide. The minimum length of a spliced mRNA would be 825 bases, and the minimum length of an unspliced pre mRNA (containing the complete intron sequence) would be 1619 bases. The unspliced RNAs detected are therefore carrying up to ca 4400 bases of additional information. It has been assumed that the most strongly hybridising transcripts (detected by the exon 1 specific probe) ranging in size from 1.95 - 3.5 Kb represent the mature mRNA(s); however this assumption may not be valid and in the absence of a cell

free translation system for mt RNAs (Chapter 3) cannot be tested easily. An alternative approach could be to isolate polysomes, dissociate the mRNA, electrophorese on an agarose gel, blot to nitrocellulose and hybridise with a maize COII specific DNA probe. Many yeast mt RNAs carry additional information for example the mRNAs encoding ATPase 9 and apocytochrome b have 5' leaders of ca 500 and ca 940 bases respectively (Grivell et al 1982).

4.3.6.6 COII transcripts in N C S and T type mitochondria

In C and T type mitochondria the sequences 3' and 5' of the COII gene are known to have undergone alterations compared to the normal (N)sequences (A. Dawson pers. commun.) and thus a preliminary investigation into the effects of these rearrangements on transcription and processing of transcripts was made. The COII polypeptide is known to be produced in normal amounts in all four cytoplasms in four day old etiolated coleoptiles (C.J. Leaver pers. commun.), hence any changes in the transcription of the gene may not be manifested in the final protein products. However studying the effects of a DNA rearrangement on COII transcription may be useful as a model for studying the expression of the altered polypeptide thought to be involved in cytoplasmic male sterility (Chapter 1 section 1.8.2).

The rearrangements around the COII gene in the mt DNA of male sterile cytoplasms have not been fully characterised; the present state of knowledge can be summarised thus (A.J. Dawson pers. commun.) [NB Studies on the rearrangement of COII sequences have been limited to the 'main' copy of COII, ie that lying on the 2.4 Kb EcoRI fragment.]:

1) No rearrangements around COII occur within ca 100 bp 5' to the AUG codon suggested by Fox and Leaver (1981) to represent the start

of the polypeptide coding sequence or ca 700 bp to 3' of the protein coding sequence, since pZmE1 hybridises to identically sized EcoRI fragments from DNA of N C S and T type cytoplasms.

2) Hybridisation of pZmE1 to HindIII digested N C S and T mt DNA identifies two fragments in each case. In N and S type mt DNA these fragments are 5.1 and 1.45 Kb. The larger fragment contains the N terminus of the gene. In C and T type mt DNA the 1.45 Kb fragment is conserved but the 5.1 Kb fragment is replaced by fragments of 3.7 and 2.8 Kb respectively. These changes are not due to restriction site changes since when the 5.1 Kb HindIII fragment is probed against a HindIII digest of T mt DNA only one fragment of 2.8 Kb hybridises (rather than two of 2.8 and 2.3 expected from the introduction of a novel HindIII site). Thus the COII gene in cms C and T mt DNA has undergone rearrangements 5' to the COII gene in C and T-type mt DNA.

Nucleotide sequencing has shown that the divergence in T occurs at 627 bp 5' to the COII AUG initiation codon. The point of divergence in C mt DNA and the nature of any rearrangements downstream from COII in C and T have not been investigated. Since the start point of the COII primary transcript in N mt DNA is not known, it is impossible to draw any conclusions on the possible effect of the DNA rearrangement in T on transcription of the COII gene. From the results presented in Fig. 4.15 it is clear that COII transcripts differ markedly from those found in N C and S mt RNA, both in abundant and nonabundant forms. Within the resolving power of the gel system there are no differences between N and T among the low molecular weight abundant transcripts, although there are seemingly some differences between N and T in the high molecular weight species. Hybridisation with the intron-specific probe should highlight these differences. Any

repeated COII sequences contributing to the overall pattern of transcription (section 4.3.6.2) would considerably complicate the picture, but the existence of these additional copies in C S and T mt DNA has not been investigated. Identification of the COII primary transcripts in N C S and T mt RNA together with more detailed analysis of the DNA rearrangements should go some way towards explaining the differences seen.

It is possible that in cms C S and T mt DNA, the COII gene is present in different genomic 'environments' which alter its expression. Effects on transcription/initiation would be manifested as differences in the primary transcript whereas effects on subsequent processing would be seen among the other transcripts.

4.4 TRANSCRIPT ANALYSIS OF THE MAIZE MITOCHONDRIAL COI GENE

4.4.1 The maize COI gene

The maize mitochondrial gene encoding cytochrome oxidase subunit I was identified by heterologous hybridisation with the corresponding yeast and bovine genes. The maize COI gene has a protein coding sequence of 1581 bp and is not split by introns (Isaac et al 1984) unlike the homologous gene in S. cerevisiae which contains a strain-dependent number of introns, for example strain D273-10B contains two introns (Nobrega and Tzagoloff 1980) whereas KL14-4A contains five introns (Lazowska et al 1980).

Transcript analysis of this gene was undertaken for the following reasons:

1) to determine whether the putative COI gene is actively expressed in the maize mitochondrion

2) to identify the transcripts of the maize mt COI region and map them to the genome.

The M13mp8 clones containing defined restriction fragments of the maize COI region used in this study were kindly provided by Dr. P.G. Isaac.

4.4.2 Transcription of the maize mt COI gene

To identify transcripts of the maize mt COI gene, mt RNA was prepared from etiolated shoots, electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, blotted to nitrocellulose and hybridised with a ^{32}P labelled DNA probe completely internal to the protein coding sequence. The probe used was a 360 bp MspI fragment spanning nucleotides 250 - 610 (clone M2A3)(Fig. 4.17).

The transcript(s) identified by this probe are shown in Fig. 4.18. The probe hybridises to a broad band of mt RNA of approximately 2.3 Kb. A longer electrophoretic run revealed that this broad band in fact represented two transcripts separated by approximately 100 bases (sized from markers not shown in Fig. 4.18). Prolonged exposure of the autoradiograph revealed the presence of a weakly hybridising band of approximately 2000 bp.

4.4.3 Mapping transcripts of the COI region

In order to identify in more detail the 5' and 3' ends of the two COI transcripts with respect to the protein coding sequence, identical RNA blots were hybridised with ^{32}P labelled probes specific for certain regions of the gene and its flanking sequences (see Fig. 4.19 a for the location of these probes). These sub-fragments extended approximately 810 bp 5' to the AUG codon, and 1,280 bp 3'

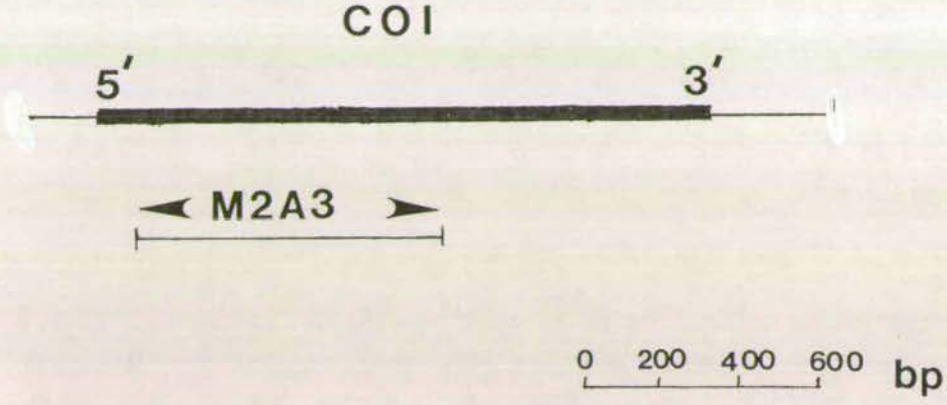


Fig 4.17 Origin of the maize COI specific clone M2A3 used to characterise COI transcripts.

Thickened bar represents the protein coding sequence. Arrows delimit the 360 bp MspI fragment cloned in M13 (clone M2A3).

Fig. 4.18 Transcripts of the maize COI gene

Mt RNA was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with ^{32}P labelled clone M2A3 (COI specific DNA probe).

Track a) Size markers: cowpea chlorotic mottle virus RNA, E. coli RNA, maize mt RNA and tobacco mosaic virus RNA. These were excised from the gel and stained separately.

Track b) Autoradiogram of RNA hybridising to the maize COI specific DNA probe.

Track c) Longer electrophoretic run of mt RNA to resolve the two major COI transcripts.

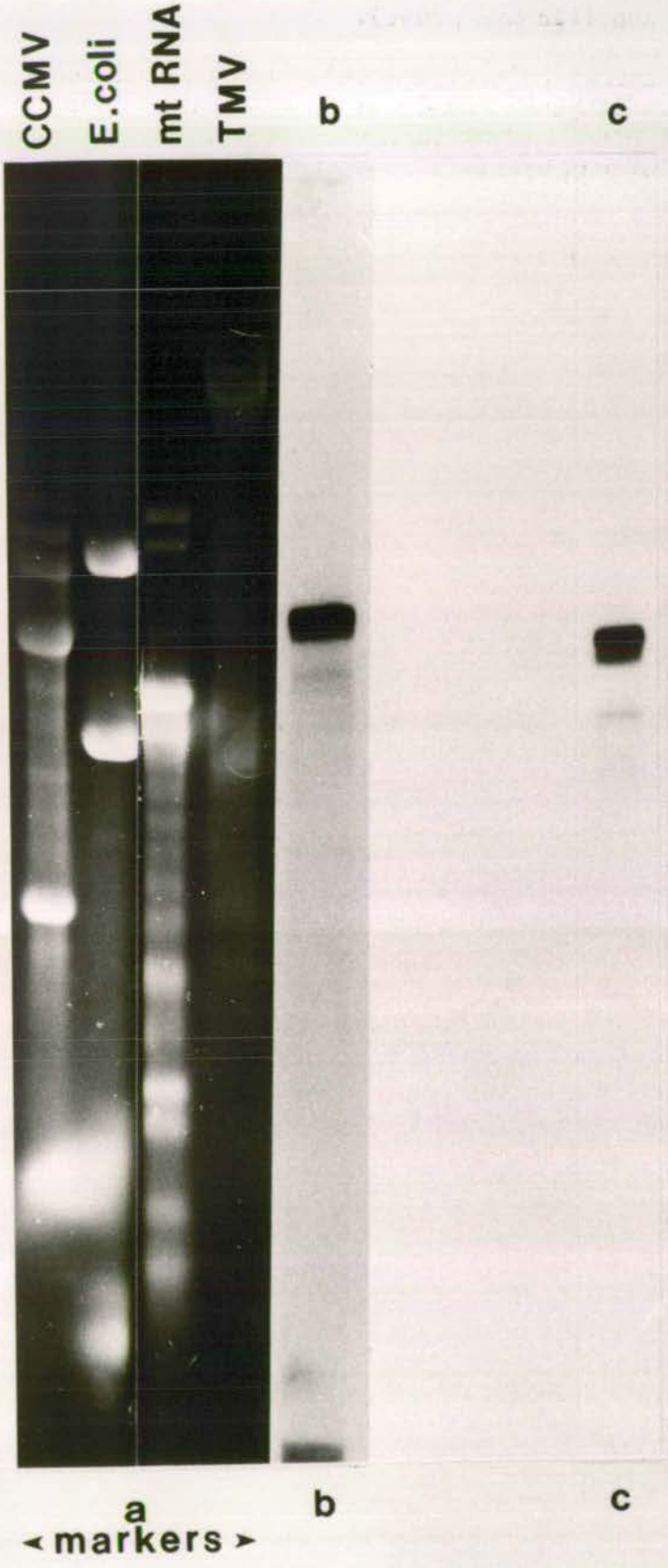


Fig. 4.18

to the UAG codon. 50µg of mt RNA was loaded into a single well, electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel and transferred to nitrocellulose. Parallel strips were hybridised with ^{32}P labelled DNA probes 1 - 8 as outlined in Fig. 4.19b.

The results of this analysis are shown in Fig. 4.19a.

This mapping experiment showed:

1) the major transcripts of the maize COI gene are two RNAs of approximately 2,400 bases and 2,300 bases. In addition probes 5 and 6 hybridise to a number of low molecular weight RNA molecules.

2) the 3' termini of both major COI transcripts map within the same DNA fragment (probe 7). This probe spans a region 393 to 634 bp downstream from the UAG codon.

3) the two transcripts have different 5' termini. The 5' terminus of the higher molecular weight maps within clone 3. This DNA probe does not hybridise to the lower molecular weight transcript.

4.4.4 S1 Nuclease mapping the 5' termini of maize mt COI transcripts

In order to map in more detail and determine the sequences around the 5' termini of the two major COI transcripts, S1 nuclease protection analysis was undertaken. Nuclease S1 from the fungus Aspergillus oryzae (Vogt 1973) is an endodeoxyribonuclease and an endoribonuclease. The enzyme specifically hydrolyses single stranded polynucleotides whereas base paired nucleotides are not hydrolysed.

If an RNA molecule is hybridised to the coding strand of a DNA restriction fragment which spans the 5' end of the transcripts, and the hybrid is digested with nuclease S1, the single stranded tails of the DNA:RNA hybrid are digested to yield a DNA:RNA hybrid of length

Fig. 4.19 Mapping COI transcripts to the mitochondrial genome

4.19 a Maize mt RNA was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with probes 1 - 8 as shown in Fig. 4.19 b.
Maize mt RNA electrophoresed in a parallel track, excised and stained with ethidium bromide provided markers (track m).

4.19 b Origin of the probes used to map the COI transcripts in relation to the COI protein coding sequence (thickened bar). Vertical bars delimit the probes used.
A + sign against a particular probe indicates hybridisation.

Fig. 4.19a

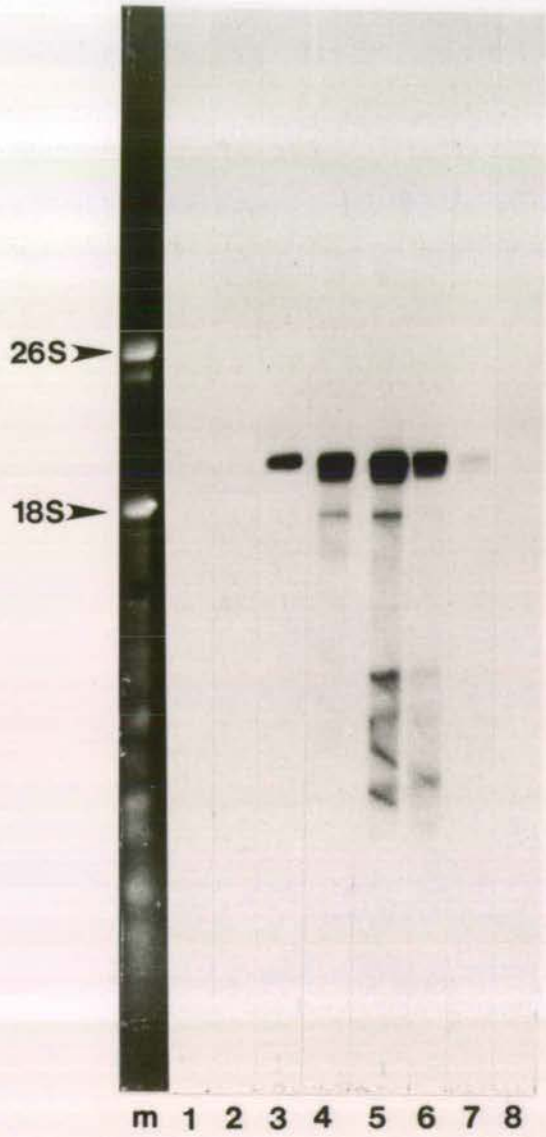
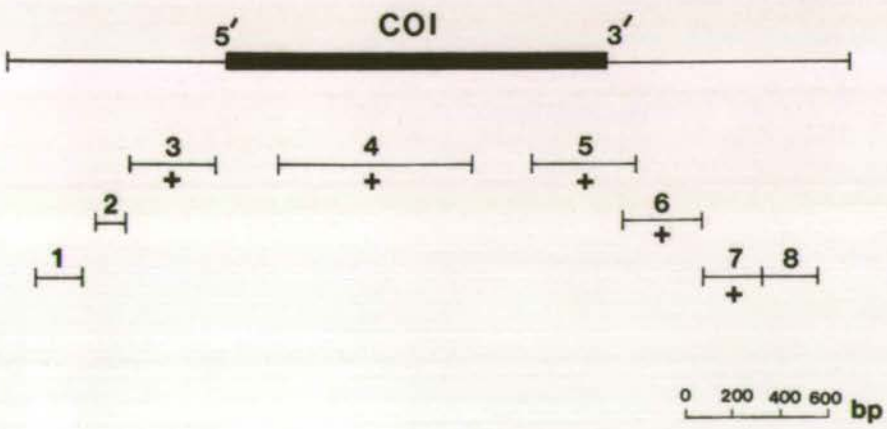


Fig. 4.19b



determined by the position of the 5' of the transcript and the restriction enzyme used. The size of the S1 resistant DNA can be estimated by electrophoresis on polyacrylamide sequencing gels with suitable markers. All S1 nuclease analysis was carried out with double stranded DNA probes which were hybridised to the RNA in high concentration formamide buffer, thus favouring DNA:RNA hybridisation over DNA:DNA hybridisation (Casey and Davidson 1977).

4.4.4.1 DNA probes used for S1 nuclease mapping COI transcripts

Two types of DNA probe were used for S1 nuclease mapping:

1) M13 single stranded clones labelled by second strand synthesis using Klenow fragment of DNA polymerase I. Only one DNA strand is labelled in this reaction, and this is chosen to be complementary to the RNA.

2) DNA restriction fragments labelled at their 5' termini using T4 polynucleotide kinase and $\gamma\ ^{32}P$ dATP

The advantage of M13 probes was the ease with which they were ^{32}P-labelled by second strand synthesis to a high specific activity ($5 - 10 \times 10^6$ dpm/μg). Also since the only strand that is labelled is that complementary to the RNA, any hybridisation with transcripts from the opposite strand will be undetected. For an unknown reason the background observed with these probes is higher than that when terminally labelled restriction fragments are used (see for example Humphries et al 1982). This may be due to steric effects of the M13 DNA on S1 digestion (leading to over or under digestion). DNA restriction fragments labelled at their 5' ends suffer the potential disadvantage that if the strands are not separated on an alkaline agarose gel prior to hybridisation with the RNA, and if both strands are transcribed, a dual set of DNA:RNA hybrids will be formed.

Both types of labelled probe were used for S1 nuclease analysis of COI transcripts. The probes used were as follows:

M13 mp8 probes labelled by second strand synthesis

1) Clone H2C2. This 354 bp HaeIII clone hybridises to the higher molecular weight COI transcript but not to the lower one. This probe was therefore used to map the 5' terminus of the higher molecular weight transcript.

2) Clone M3A3. This 760 bp MspI clone spans the start of both COI transcripts.

Double stranded DNA restriction fragment labelled at its 5' terminus

1) A 760 bp MspI fragment identical to that contained in clone M3A3. In order to isolate this fragment, plasmid pBN6601 (a 3.9 Kb BamH1/EcoRI fragment of mt DNA containing the entire COI gene and 2319 bp of flanking sequences cloned in pAT153 Isaac et al 1984) was digested with MspI (Fig. 4.20 a). The 760 bp MspI fragment was eluted from the gel, the 5' terminal phosphates removed using CIP, and labelled at the 5' termini using γ^{32}PdATP and T4 polynucleotide kinase. The purity of the labelled fragment was checked by electrophoresis through a 4% (w/v) polyacrylamide gel followed by autoradiography (Fig 4.20 b). The genomic locations of the probes used for S1 nuclease mapping is shown in Fig. 4.21.

4.4.4.2 Nuclease S1 mapping COI transcripts with M13 clone H2C2

M13mp8 clone H2C2 (approximately 1µg of single stranded DNA) was ^{32}P labelled by second strand synthesis to a specific activity of ca 5 - 10 x 10^6 dpm/µg. The labelled DNA was divided into three equal aliquots and ethanol precipitated with other nucleic acids as follows:

1) DNA only control - indicates the amount of renaturation that has occurred between the two separated DNA strands. This can occur if

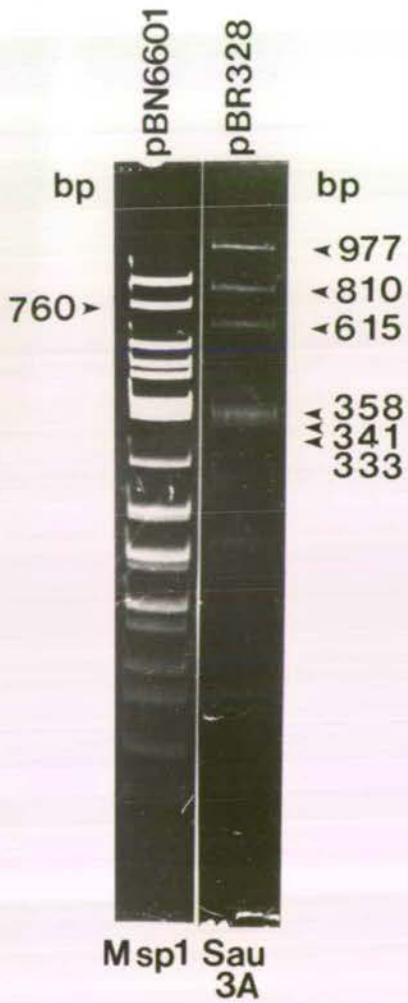


Fig. 4.20a

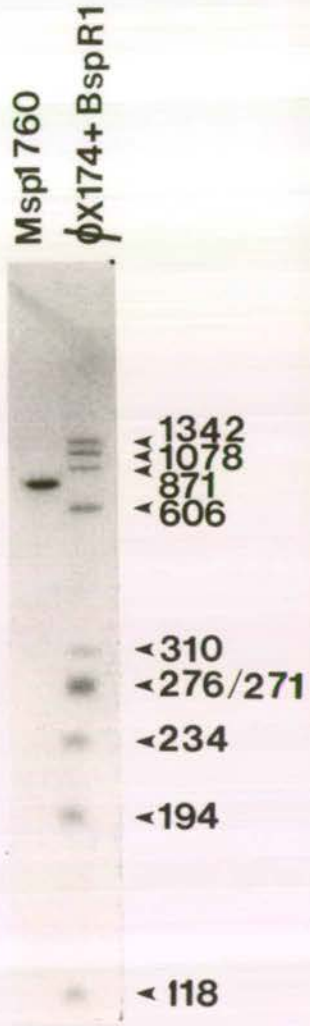


Fig. 4.20b

Fig 4.20 Purification and 5' labelling of a 760 b_p MspI COI fragment

4.20a pBN6601 was digested with MspI and electrophoresed on a 6% (w/v) polyacrylamide gel. Markers were provided by pBR328 digested with Sau 3A.

4.20b The MspI 760 b_p fragment was ^{32}P labelled at its 5' termini using T4 polynucleotide kinase and γ ^{32}P dATP. The purity of the labelled fragment was assessed by electrophoresis on a 4% (w/v) polyacrylamide gel followed by autoradiography.

ϕ X174 digested with Bsp RI similarly ^{32}P labelled provided markers.

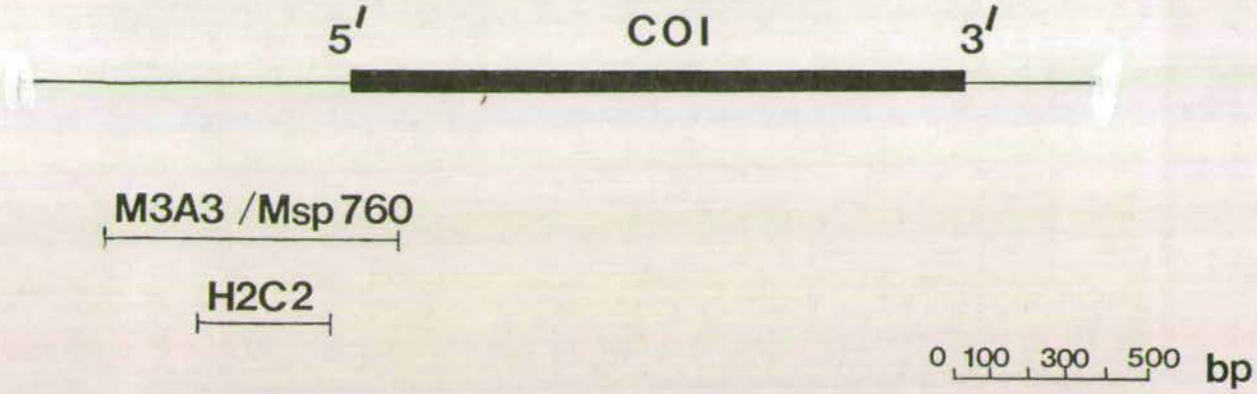


Fig.4.21

Fig. 4.21 Locations of the COI probes used for S1 mapping the 5' termini of the maize mt COI transcripts

Protein coding sequence is represented by thickened bar. Vertical lines delimit the 5' and 3' ends of the DNA probes used.

quenching with ice cold S1 buffer after the hybridisation reaction is not sufficiently rapid (Sharp et al 1980).

2) DNA + tRNA control - ^{32}P labelled DNA was ethanol precipitated with 15 - 25µg E. coli tRNA. This control is a measure of any non-specific hybridisation of the DNA with RNA.

3) DNA + mt RNA - ^{32}P labelled DNA was ethanol precipitated with 15 - 25µg maize mt RNA. This formed the actual experimental hybridisation.

The co-precipitated nucleic acids were resuspended in hybridisation buffer (Chapter 2) and submerged in a water bath at 83°C for 15 minutes to denature the DNA strands. The tubes were rapidly transferred to a water bath set at the required temperature where they were incubated submerged for 3 hours. The required hybridisation temperature was calculated from the formula presented in Chapter 2, section 2.2.5.3. Clone H2C2 insert has a (G + C) content of 41.4%. Thus the DNA:DNA T_m of this fragment in the hybridisation buffer is ca 35°C. A range of hybridisation temperatures were chosen which were T_mDNA: DNA +5°C, +8°C and +11°C. All three hybridisation temperatures gave identical results but the strongest signal with least background was obtained at hybridisation temperatures of T_m DNA:DNA +8°C (= 43°C in this case).

The DNA:RNA hybrids were treated with S1 nuclease at 400 U/ml at 37°C for 20 minutes. The S1 nuclease concentration was arbitrarily chosen as the effect of a range of S1 nuclease concentrations on quantity and electrophoretic mobility of DNA:RNA hybrids was not investigated. The concentration of S1 nuclease added is apparently not critical as a tenfold increase does not affect the mobility or

quantity of DNA:RNA hybrids formed in protective experiments with adenovirus transcripts (Sharp et al 1980, John Clark pers. commun.).

After phenol extraction the DNA:RNA hybrids were denatured and electrophoresed on 6% (w/v) polyacrylamide/8M urea sequencing gels in parallel with the products of an M13 sequencing reaction.

Fig. 4.22 shows the S1 nuclease resistant DNA:RNA hybrid formed with probe H2C2 (arrowed). The length of the protected DNA fragment was calculated by determining with which base of the M13 sequencing reaction it co-migrated. (A sequencing reaction using Clone H2C2 template DNA identical to that used in the S1 reaction was used as a marker.) Accounting for the 'smile' of the gel, the S1 nuclease protected DNA fragment co-migrates approximately with the G residue denoted by an asterisk in Fig. 4.22. This of course does not mark the 5' terminus of the high molecular weight transcript, since the length of the marker track must be adjusted for the additional sequence present between the 5' end of the pentadecamer primer and the cloning site used in the polyclonal linker (SmaI). In this case therefore, the 5' terminus of the transcript maps approximately 64 bp upstream from this G residue marked G*. The length of the protected hybrid is therefore 104 bp from the HaeIII site.

Hence the DNA sequence around the start of the higher molecular weight transcript is

5' CATAAGTAATC 3'
*

where * represents the approximate 5' end of the higher molecular weight transcript. The distance between the 5' end of the COI higher molecular weight transcript and the AUG codon is therefore approximately 156 bp.

Fig. 4.22 S1 nuclease mapping the 5' end of the maize COI higher molecular weight transcript

M13 clone H2C2 which contains the start point of the higher molecular weight transcript was ^{32}P labelled by second strand synthesis and hybridised with 25µg maize mt RNA. Hybrids were digested with S1 nuclease (400 U/ml) and electrophoresed through a 6% (w/v) polyacrylamide/8M urea sequencing gel.

Track a) H2C2 plus 25µg mt RNA
Track b) H2C2 plus 25µg yeast tRNA

The size of the S1 protected DNA:RNA hybrids were estimated by alignment with an M13 sequencing ladder of clone H2C2.

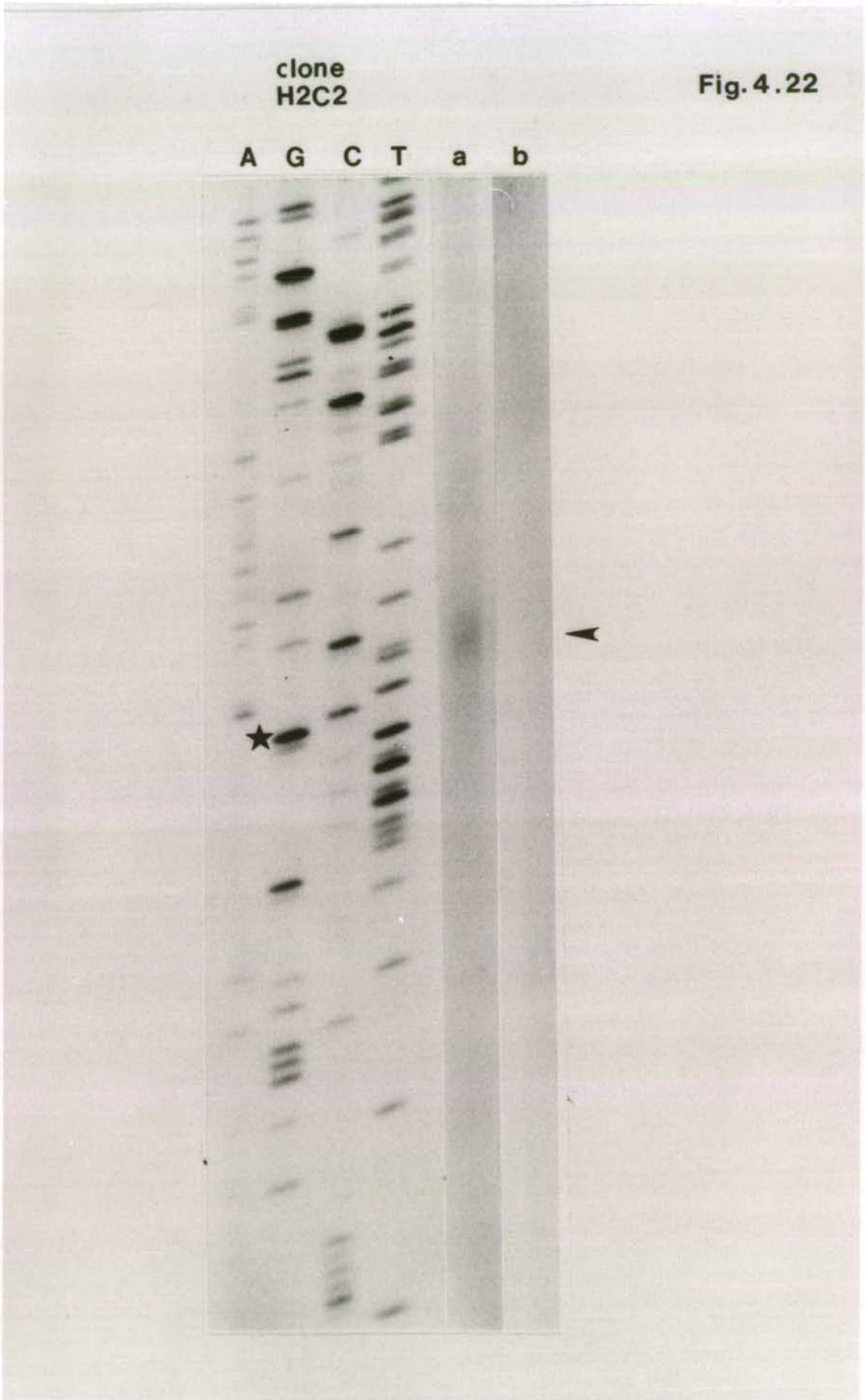
The most strongly protected band is marked ◄

★ marks the G residue with which the strongest S1 protected band co-migrates.

Note that the marker track was not immediately adjacent to tracks a and b, and thus the 'smile' of the gel has been taken into account when sizing the protected fragments.

clone
H2C2

Fig. 4.22



4.4.4.3 Nuclease S1 mapping COI transcripts with M13 clone M3A3

M13mp8 clone M3A3 (approx. 1µg) was labelled by second strand synthesis to a specific activity of $> 10^6$ dpm/ug. The DNA was divided into three equal aliquots and precipitated with DNA, tRNA and mt RNA as described in section 4.4.4.2. The co-precipitated nucleic acids were resuspended in hybridisation buffer and denatured as before. The tubes were rapidly transferred to a water bath at the hybridisation temperature of 43°C. This was calculated to be 8°C above the T_m DNA:DNA of clone M3A3 insert (43% [G + C]). Hybridisation was allowed to proceed for three hours, after which the DNA:RNA hybrids were treated exactly as described in section 4.4.4.2.

Fig. 4.23 shows the S1 nuclease protected DNA:RNA hybrids formed with probe M3A3. Clearly both the low molecular weight and the high molecular weight transcripts have formed DNA:RNA hybrids of different length. In order to determine accurately the 5' start points of these two transcripts, an electrophoretic run of six hours was carried out. Markers were provided from a chain termination sequencing reaction of M3A3. The major protected DNA fragment generated by hybridisation with lower molecular weight transcript (denoted by large closed arrow) migrates with a thymidine residue denoted by an open arrow on the autoradiogram. This T residue is 141 bp from the MspI cloning site (AccI). Hence the length of the protected fragment is 141 bp + 54 bp contributed by the pentadecamer primer and the polyclonal linker up to the AccI site. Thus the total length of the protected fragment is 195 bp, and the sequence around the start of the lower molecular weight transcript is

5' CCTTCA*TTCTTT 3'

where * represents the approximate 5' end of the lower molecular weight

Fig. 4.23 S1 nuclease mapping the 5' ends of the maize mt COI transcripts

Clone M3A3 was ^{32}P labelled by second strand synthesis and hybridised with 2.5µg mt RNA (track b) and 25µg yeast tRNA (track a). Hybrids were digested with S1 nuclease (400 U/ml) and electrophoresed through a 6% (w/v) polyacrylamide/8M urea sequencing gel.

The size of the S1 nuclease protected DNA hybrids were estimated by alignment with an M13 sequencing ladder of clone M3A3.

S1 protected hybrids are denoted by closed arrows.
Open arrow denotes T residue with which lower molecular weight transcript protected hybrid migrates.

Fig. 4.23

transcript. The distance between the 5' end of the lower molecular weight transcript and the AUG codon is therefore approximately 61 bp.

The long electrophoretic run (6 hrs) did not sufficiently resolve the DNA marker fragments to allow an approximate estimate of the size of the larger protected fragment. However it is clear that both the larger and the smaller protected fragments have resolved into groups of bands. The higher molecular weight transcript seems to generate three protected fragments and the lower molecular weight transcript generates two.

4.4.4.4 Nuclease S1 mapping COI transcripts with the 5' ^{32}P labelled 760 MspI fragment

This experiment was carried out in the Department of Biochemistry, University of Amsterdam. The 5' ^{32}P labelled 760 bp MspI restriction fragment was hybridised with mt RNA exactly as described for clone M3A3 in section 4.4.4.3. DNA only and DNA + tRNA controls were also set up. The DNA:RNA hybrids were treated with S1 nuclease as before except that the final concentration of S1 was 100 U/ml instead of 400 U/ml. The protected hybrids were electrophoresed on a 6% (w/v) polyacrylamide/8M urea sequencing gel. A long electrophoretic run (6 hr) sufficiently resolved the DNA marker tracks to allow the approximate size of the larger protected hybrid to be determined (Fig. 4.24). Size markers were provided by a sequencing ladder of a yeast mt DNA MboI clone in M13mp8 provided by M. de Haan.

Lower molecular weight transcript:

The protected hybrids from the lower molecular weight RNA species forms a diffuse band. This band co-migrates with a C residue in the marker DNA track (denoted by *). This allowed the size of the protected fragment to be estimated at approximately 197 bp. Thus, the

Fig. 4.24 S1 nuclease mapping the 5' ends of the maize mt COI transcripts

A 760 bp MspI fragment was ^{32}P labelled at its 5' termini using T4 polynucleotide kinase and $\gamma\ ^{32}P$dATP. The labelled DNA was hybridised with 15µg mt RNA (track c) and 15µg yeast tRNA (track b). Track a represents input DNA not digested with S1 nuclease. S1 protected fragments are denoted by closed arrows. The sizes of the S1 nuclease protected fragments were estimated by alignment with an M13 sequencing ladder of a yeast mt DNA M13 clone.

Open stars mark the residues of the sequencing ladder with which the protected fragments co-migrate.

The open arrow denotes a faint high molecular weight S1 nuclease protected band.

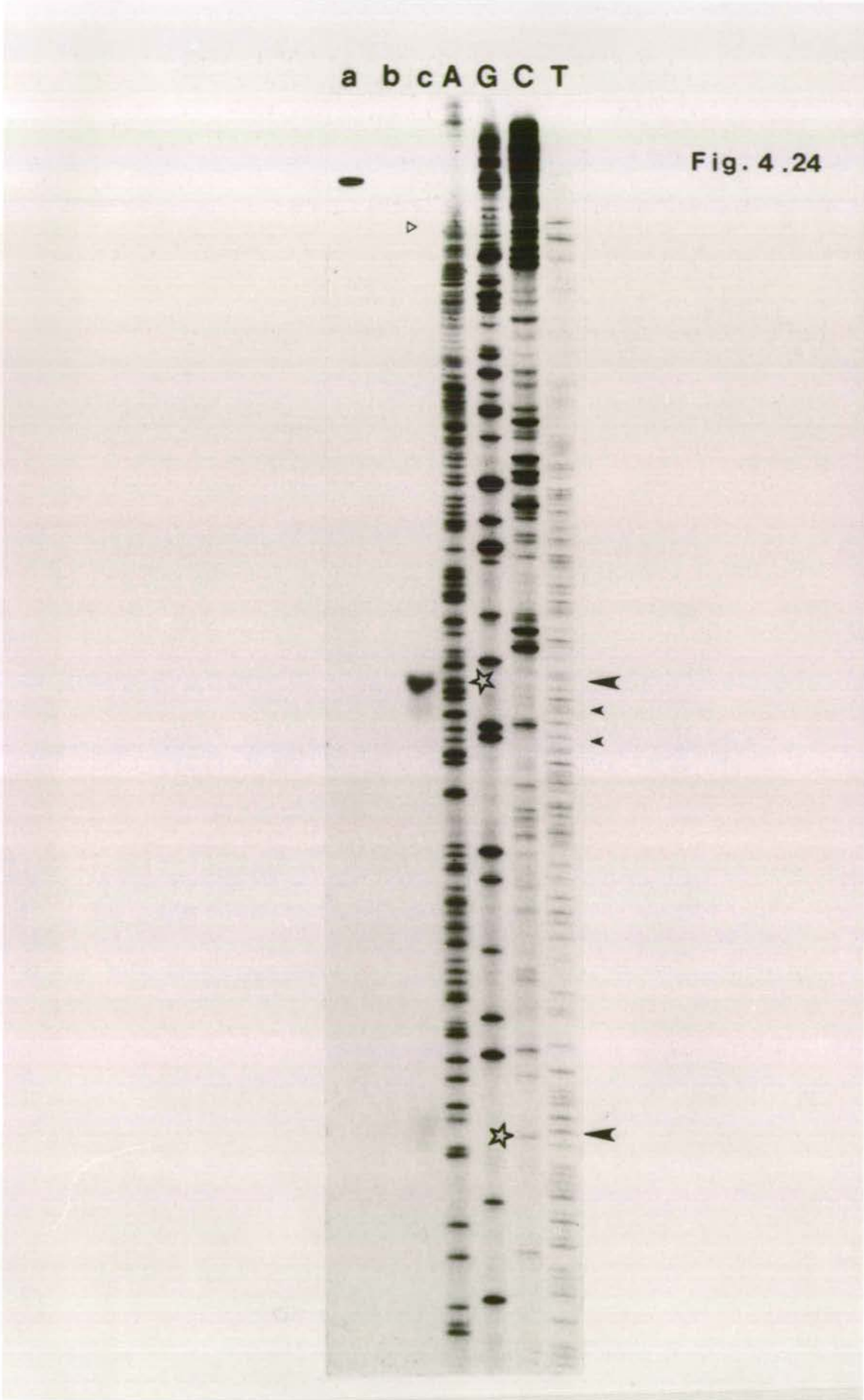


Fig. 4.24

sequence around the start of the lower molecular weight transcript is

5' CATTC*TTT 3'

where * represents the predicted 5' end of the transcript, which places the 5' end of the lower molecular weight transcript 57 bp from the AUG initiation codon.

Higher molecular weight transcript

The DNA fragment protected by the higher molecular weight transcript is again resolved into three discrete bands by this probe, separated by approximately 10 - 20 bases. The significance is not clear since multiple bands were not observed with the probe H2C2; however this may be simply due to a difference in specific activity of the probes used and the length of exposure to X-ray film. These groups of bands are discussed later (section 4.4.5) and only the uppermost is considered here. The protected fragment co-migrates with an A residue (marked *) and the length of the protected fragment is approximately 287 bp. The sequence around the higher molecular weight transcript is

5' CATAAGTAA*TCC 3'

where * denotes 5' end of the transcript, which is ca 152 bp from the COI AUG initiation codon.

In addition to these protected hybrids a third very high molecular weight fragment is detected (not reproduced in Fig. 4.24 but denoted by ▷), of slightly lower molecular weight than the input 760 bp MspI DNA restriction fragment. It is not observed in the DNA:tRNA control and thus cannot represent a DNA:DNA renaturation intermediate (Sharp et al 1980). Since it was not observed when single stranded M13 probes were used for S1 mapping, it is possible that it represents a hybrid formed with a transcript initiating from the opposite DNA

strand.

4.4.4.5 Summary of nuclease S1 mapping data

The 5' termini of the two major COI transcripts have been approximately mapped to the genome and the sequences around the start of the two transcripts identified.

1) The 5' end of the higher molecular weight transcript maps within the sequence

$$5'\ CATA\overset{*}{A}G\overset{*}{T}AATC\ 3' \overset{149\ bp}{\ldots\ldots}\ AUG$$

where * denote the 5' ends as determined in two separate experiments using different probes labelled by different means.

2) The 5' end of the lower molecular weight transcript maps within the sequence

$$5'\ CCTTC\overset{*}{A}TTC\overset{*}{T}TT\ 3' \overset{54\ bp}{\ldots\ldots}\ AUG$$

where * denote the 5' ends as determined in two separate experiments using different probes labelled by different means.

4.4.5 COI transcripts in N C S and T mt RNA

Rearrangements of the COI gene in male sterile maize lines have not been extensively investigated. Those that are known to exist are discussed in section 4.4.6.6.

Mt RNA prepared from N (nuclear genotype 3541), N C S and T (all nuclear genotypes WF9) was electrophoresed on a 1.3% (w/v) agarose/ 7% (v/v) formaldehyde gel, blotted to nitrocellulose and hybridised with a probe specific for the COI protein coding sequence (clone M2A3) [Fig. 4.25].

The two major transcripts hybridising to the internal COI gene probe are identical in all four cytoplasms, and in normal cytoplasms from

Fig. 4.25

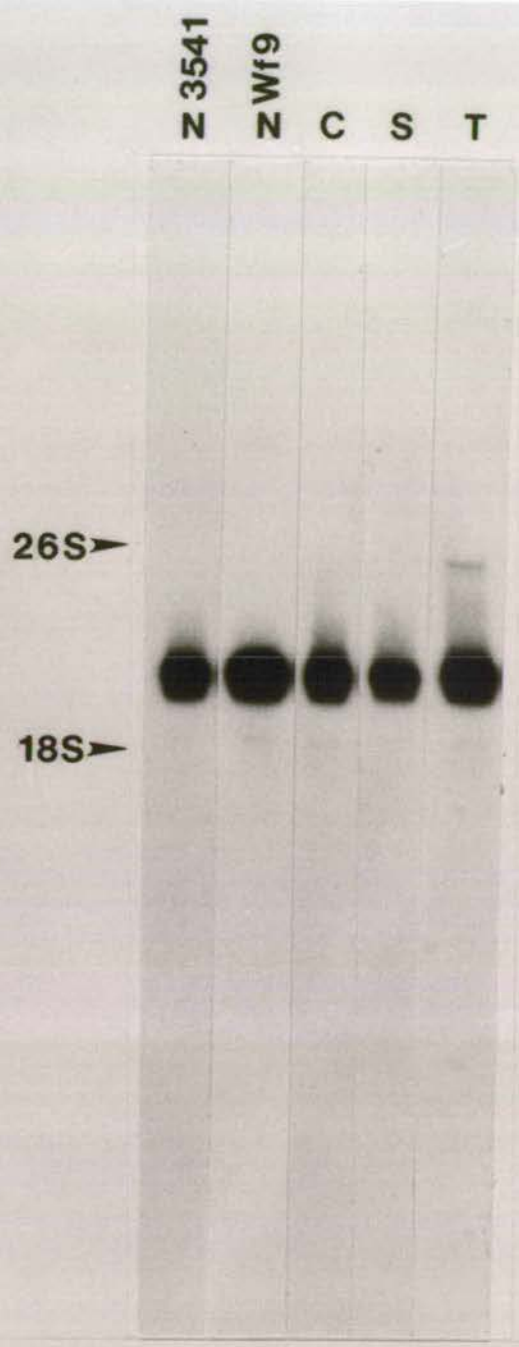


Fig. 4.25 COI transcripts in N, C, S, and T type mt RNA

Mt RNA was isolated from N (nuclear background 3541) and N, C, S,and T (all nuclear background WF9) type mitochondria, electrophoresed on a 1.3% (w/v) agarose /7% (v/v) formaldehyde gel, blotted to nitrocellulose and hybridised with a COI specific M13 clone (M2A3) ^{32}P labelled by second strand synthesis. Markers were obtained by staining the nitrocellulose with ethidium bromide to visualise the rRNAs.

10 µg of each type of RNA was loaded in each track.

two different nuclear backgrounds. Prolonged exposure of the autoradiograph reveals some changes in the minor bands detected in section 4.4.2.

4.4.6 Discussion

4.4.6.1 Transcripts of the maize COI gene

Transcript analysis of the COI gene has shown that the gene is probably actively transcribed in the maize mitochondrion. The maize COI gene is seemingly transcribed in a much less complex manner than COII. Northern blotting has identified two strongly hybridising transcripts of approximately 2.4 and 2.3 Kb. Prolonged exposure of the autoradiograph reveals another very faintly hybridising band of approximately 2.0Kb. The significance of this minor transcript is not clear. It is possible that it results from transcription of the non-coding COI strand. Although this is unlikely because strand specific probes were used in all these experiments, subsequent investigations have shown (section 4.5.2) that secondary hybridisations between the separated strands of the M13 DNA can occur. Thus the unlabelled template DNA may hybridise to transcripts of the non-coding strand, and the ^{32}P labelled M13 DNA resulting from second strand synthesis hybridises to its complement. In this way transcripts from the non-coding strand which should not be detected with strand specific probes appear as faintly hybridising bands superimposed on the strongly hybridising transcripts of the 'coding strand'. A non-coding strand origin for these minor transcripts could easily be investigated using M13 probes specific for that strand. Any transcripts from the COI 'non-coding' strand would appear as strongly hybridising RNAs superimposed on the faintly hybridising COI transcripts. It has been assumed that since these

two transcripts are the only major RNA molecules hybridising to the gene specific probe one or both must represent the mature mRNA(s). In the absence of a cell free translation system (Chapter 3) this cannot be proven, other than by isolation of polysomes and characterisation of the COI mRNA.

Both major COI transcripts are far longer than required to specify the corresponding polypeptide (1581 bp). This means both transcripts (ca 2.4 and 2.3 Kb) carry 700 and 600 bases of additional sequence information respectively. Mapping experiments have shown that the vast majority of this additional sequence is carried at the 3' end of the transcript.

In S. cerevisiae the transcripts of the COI gene are of course more complex since up to nine introns in strain KL14-4A or seven in strain D273-10B (Hensgens et al 1983, Bonitz et al 1980) must be spliced from the precursor RNA. The mature mRNA of the S. cerevisiae COI gene is approximately 2,100 nucleotides long (Hensgens et al 1983) containing 630 bases of additional information. However, in contrast to the maize COI transcript, most of this additional sequence is in the form of a long 5' leader, transcription being initiated 540 bp 5' from the AUG start codon (Osinga et al 1984).

Additional information present in most yeast mt transcripts generally takes the form of a long 5' leader rather than an extensive 3' tail. For example the S. cerevisiae COB and ATPase 9 mature mRNAs have 5' leader sequences of ca 940 and 500 nucleotides respectively, while their 3' trailing sequences are ca 100 and 70 nucleotides respectively (Hensgens et al 1979).

The exception to this is the yeast COIII mature mRNA which contains 2940 bases of additional sequence. S1 nuclease mapping has shown that the mRNA has a 5' leader of 490 bases and a 3' extension of 2450 bases (in strain D 273 - 10B, Thalenfeld et al 1983). Although COIII is known to be co-transcribed with $tRNA^{Phe}$ and $tRNA^{Val}$ this additional sequence is not part of a precursor since these polycistronic species are processed too rapidly to be detected in Northern blots. Thalenfeld et al (1983) argue that since in S. cerevisiae strain LL20 the 3' trailing region is reduced by 2 K bases it cannot play any major role in the expression of the yeast COIII gene.

An alternative explanation for the long 3' trailing sequence of the maize COI transcript is that it forms a polycistronic mRNA with coding sequences downstream of the COI gene. However no open reading frames or tRNA-like structures have been found in this region (P.G. Isaacs pers. commun.).

Recently the COII transcripts in wheat have been mapped and shown to have 5' leaders of 200 and 3' extensions of 500 bases (Bonen et al 1984). It is possible therefore that long 3' tails may be a characteristic of higher plant mt transcripts.

Although the 3' termini of the COI transcripts have not been mapped using S1 nuclease, they are known to terminate between 393 and 634 bases from the UAG codon. The difference in the 5' position of the two COI transcripts (ca 95 bases) is probably sufficient to account for the size difference of the two major transcripts and therefore they are likely to share a common termination site.

4.4.6.2 Origin of two COI transcripts

Several models can be proposed to account for the presence of two major COI transcripts.

1) There are two or more copies of the COI gene in maize mt DNA, giving rise to transcripts of slightly different molecular weights.

2) There are two separate transcription initiation events generating two transcripts differing by ca 95 bases.

3) The higher molecular weight transcript represents the primary transcript which is then processed to generate the lower molecular weight transcript.

4) Neither of these transcripts represents the primary transcript, but are both generated by very rapid processing of a primary transcript which is not detectable by Northern blotting. It should be relatively easy to distinguish between these models.

There is no evidence for more than one copy of the COI gene in N type mt DNA (P.G. Isaac, pers. commun.).

Two separate transcription initiation events giving rise to the two transcripts implies that the two initiation points have sequence homology if acting as promoters for a single RNA polymerase (as is thought to be the case in the yeast mitochondria, Tabak et al 1983). Conversely processing of the higher molecular weight transcript to generate the lower molecular weight RNA would not necessarily predict any sequence homology, since different enzymes are likely to be involved in transcription initiation and processing.

The sequences around the 5' termini of the COI transcripts are discussed in the following section (4.4.6.4).

The primary transcript of the maize COI could be detected in one of two ways:-

1) Using the guanylyl transferase assay ["Capping" (Edwards et al 1983)]. As outlined in section 4.3.6.2, this assay distinguishes between transcripts arising from a transcription initiation event and those arising from processing of primary transcripts. The enzyme [which has been purified from Vaccinia virions, HeLa cells (Shuman and Hurwitz 1982) and calf thymus (Nistukawa and Chambon 1982)] is involved in capping eukaryotic and viral mRNAs (reviewed by Shuman and Hurwitz 1982). It catalyses the transfer of GMP from GTP to the 5' terminus of RNA to form the cap structure pGpppX. Only RNAs containing 5' di- or tri-phosphate end can act as substrates for the enzyme. RNAs containing a 5' monophosphate or 5' OH terminus are not suitable substrates. Since the only known mechanism for the production of RNAs with 5' di- or tri-phosphate is transcription initiation, then the enzyme uniquely labels RNA molecules which are primary transcripts.

Edwards et al (1983) have reported that in the capping assay, where there are many potential transcription initiation sites low abundance mRNAs need to be enriched in some way in order to characterise capped RNAs. In yeast any mt mRNA can be enriched by isolating it from a suitable petite yeast strain where particular regions of the mt genome have become amplified. In maize mitochondria an obvious way of enriching for an mRNA is by "hybrid release" (Goldberg et al 1979). The transcripts specific for COI generated by hybrid release could then be capped in vitro with guanylyl transferase and sequenced directly by a modification (Peattie 1979) of the Maxam and Gilbert procedure. Comparison of the RNA sequence with the DNA sequence

upstream from COI would thus identify the start point(s) of any COI primary transcript(s).

2) As outlined in section 4.3.6.2, primary transcripts could also be identified by the use of an in vitro transcription system using purified mt RNA polymerase. Such an in vitro system has been extensively used in the characterisation of S. cerevisiae primary transcripts (see for example Tabak et al 1983). Crude preparations of mt RNA polymerase can easily be prepared according to published procedures (Levens et al 1981). The system could be primed with suitable restriction fragments of the maize COI region as templates, and the length of any "run-off" product(s) measured. In this way the start point of the primary transcripts could be determined.

4.4.6.3 DNA sequences at the 5' ends of the two major COI transcripts

S1 nuclease analysis has shown that the two major COI transcripts start at approximately 57 and 152 bases 5' from the AUG initiation codon. The DNA sequences determined by S1 nuclease mapping to be at the start points of both transcripts were examined for any similarities to each other and to the putative promoter sequence thought to be important in transcript initiation in yeast mitochondria (Osinga and Tabak 1982).

The DNA sequences around the 5' ends of both major COI transcripts show no homology to each other:

5' CATAAGTAATC 3' HMW COI transcript

5' CCTTCATTCTTT 3' LMW COI transcript

* represents the starts of transcription determined in two separate experiments.

Whilst no putative promoter sequence(s) has yet been identified in plant mitochondria, the lack of any sequence homology around the starts of these transcripts would argue against both resulting from separate transcription initiation events if plant mitochondria have a single RNA polymerase.

The DNA sequence around the 5' end of the high molecular weight transcript shows considerable homology to the putative promoter sequence in yeast mitochondria

```
                    *
5' (T)  TATAAGTA  3'     yeast nona-
   (A)                   nucleotide putative
                         promoter sequence
5' T    CATAAGTAA 3'     maize COI HMW
     *    *              transcript
```

* marks the first transcribed base, vertical lines indicate homologous sequences.

In yeast a nonanucleotide sequence is always found at points where transcription and DNA replication is initiated *in vivo*; genes lacking this 5' motif have been shown to be transcribed as part of a polycistronic precursor (reviewed in Tabak *et al* 1983, Edwards *et al* 1983, Osinga *et al* 1984). This nonanucleotide sequence is thought to be an essential part of a yeast mt promoter in view of its location, and the fact that when it is absent in petites, gene transcription is not correctly initiated (Thalenfeld *et al* 1983, Osinga *et al* 1982)

The DNA sequence around the maize high molecular weight COI transcript shows 7/9 or 8/9 homology to the yeast putative promoter (the first nucleotide of the yeast putative promoter is usually A but can be T; Tabak *et al* 1983). If this consensus is found upstream

from any other maize mt genes it would strongly suggest that it formed part of a maize mitochondrial promoter which is related to that of yeast. Recently the 5' end of the wheat COII gene transcript has been mapped to the genome by S1 nuclease mapping and shown to start approximately 175 (± 10) bp from the AUG initiation codon (L. Bonen pers. commun.). Comparison of the sequence around the start of the wheat COII transcript with that of the maize COI HMW transcript and the yeast putative promoter sequence reveals striking homologies:

```
wheat COII      5'  A TATAANTA   3'
                    ⋮ |||||  ||
yeast           5' (T)TATAAGTA   3'
                   (A) |||||||
                    ⋮ |||||||
maize COI HMW   5'  T CATAAGTAA  3'
```

* marks the first base of transcripts. In wheat, the start point of the COII transcript has not been accurately determined.

Comparison of the DNA sequence around the start of the maize COI HMW transcript with the only other higher plant mitochondrial RNA to be mapped to the genome reveals an even more striking homology. Beta vulgaris mitochondria contain a 1.44 Kb 'plasmid like' circular DNA molecule. This circular molecule is transcribed yielding a major RNA product of 850 bases (Munk Hansen and Marcker 1984). The sequence around the 5' end of this transcript is compared with that of the maize COI HMW transcript below

```
                **
5'  CTAAAATCATAAGTGATAT   3'   Beta vulgaris
    ||| |||||||||| ||          1440 bp plasmid
                               transcript
AGAAACTCATAAGTAATCC
    *    *                     maize COI HMW
                               transcript
```

* represents the possible first base of transcripts, vertical lines indicate identical bases.

Thus the homology around the start of these two mitochondrial transcripts in two different higher plants is 13/15. Thereafter homology drops off rapidly.

Clearly, identification of the DNA sequence around the start points of the primary transcripts of other plant mt genes is necessary to determine whether these homologies have arisen by chance or whether they represent part of a plant mitochondrial promoter.

The DNA sequence around the 5' end of the LMW COI transcript does not show any homology to the yeast nonanucleotide putative promoter sequence, nor to the DNA sequence around the 5' end of the Beta vulgaris mt plasmid transcript, nor to the wheat COII transcript. It does however show 6/12 homology to the dodecamer sequence thought to be involved in processing (as opposed to splicing) of yeast mt polycistronic transcripts (Osinga et al 1984).

```
5'  AATAATATTCTT  3'   yeast dodecamer sequence
        ||||||
5'  CCCTTCATTCTT  3'   maize COI LMW transcript
          *  *
```

In yeast mitochondria this dodecamer sequence is found 3' from nearly all protein coding genes and is thought to represent a processing point for the generation of mature transcripts from polycistronic messages (rather than as a signal for transcription termination, see Chapter 1 section 1.6.3). If this were the case in maize mitochondria, and the 6 out of 12 homology with the yeast dodecamer sequence is significant, then the lower molecular weight COI transcript could arise either from processing of the higher molecular weight transcript or from very rapid processing of an even higher molecular weight transcript not detected on Northern blots. Clearly however, identification of the COI primary transcript(s) and analysis of the

transcripts of other plant mt genes is necessary before predictions of this sort can be made with any confidence.

4.4.6.4 Possible heterogeneity of the COI transcripts

The 5' ends of both major COI transcripts do not generate sharp bands when subjected to S1 nuclease analysis. The HMW transcript appears to resolve into three groups of bands separated by ca 10 - 20 bases. The lower molecular weight transcript resolves into two groups of bands. This effect is unlikely to be due to incomplete or over digestion by S1 since the same pattern is seen over a four fold range of S1 concentrations. One possible explanation is that the 5' ends of the COI transcripts are 'ragged'.

Ragged ends have also been found for the 3' end of the S. cerevisiae ATPase 9 transcript (Thalenfeld et al 1983) and for the 5S and 18S rRNAs in wheat mitochondria (Spencer et al 1981, Schnare and Gray 1982, Spencer et al 1984). Primer extension (Ghosh et al 1978) using synthetic oligonucleotides or suitable restriction fragments would confirm whether these ragged ends of the COI transcripts detected by S1 nuclease analysis are real phenomena, and not artefacts of S1 digestion.

4.4.6.5 COI transcripts in N C S and T mitochondrial RNA

Rearrangements associated with the COI gene in cms C S and T mitochondrial DNA have not been extensively investigated (P.G. Isaac pers. commun.). However it is known that:

1) In S mt DNA there are repeated copies of the COI gene. Two HindIII fragments of 10.0 and 8.8 Kb hybridise strongly to an internal COI probe, as well as four fragments of 8.79, 6.3 (x 2) and 4.9 Kb which hybridise weakly. At least one of the copies of the COI gene

in S mt DNA is complete.

2) Sequence analysis 5' of one of the S mt DNA COI genes reveals a major rearrangement 174 bp 5' from the AUG initiation codon. Up to the point of rearrangement the sequence is identical with N mt DNA. Thereafter S mt DNA shows extensive homology with the 208 bp terminal inverted repeat of the episomal DNAs S1 and S2 (section 1.8.2.2).

3) In cms T mt DNA a rearrangement occurs 3 - 5 Kb 3' to the UAG termination codon. Hybridisation data indicates that whereas a 9.2 Kb EcoRI fragment of N mt DNA hybridises to a gene internal probe, in T a 7.2 Kb fragment is identified. More extensive mapping data suggests that the alteration occurs between 3 and 5 Kb downstream of the 3' end of the gene.

Rearrangements of COI in C mt DNA have not been investigated.

It can be concluded that since no differences were found in RNAs from N C S and T mt RNA hybridising to COI DNA probes, these rearrangements do not have any significant effect on transcription of COI. The significance of the changes detected in weakly hybridising transcripts is not known.

It is particularly striking that the rearrangement occurring 174 bp 5' from the COI initiation codon does not affect the pattern of transcripts hybridising to the COI specific probe. Since the higher molecular weight COI transcript initiates around 152 bp from the AUG initiation codon, it is likely that this RNA species actually represents the primary transcript. A primary transcript which was initiated further upstream would most likely be affected by the rearrangement occurring in S mt DNA. Alternatively if both COI major transcripts are processed from a larger primary transcript this

could still be processed correctly in S mitochondria since the processing site (assumed to be around the 5' start of the COI transcript) is not altered in S mt DNA. In this case however the secondary structure of the primary transcript would be completely different from N mt RNA, and this may affect processing. Identification of the primary COI transcript would resolve these possibilities.

4.5 TRANSCRIPT ANALYSIS OF THE MAIZE APOCYTOCHROME b GENE

4.5.1 The maize COB gene

The maize mitochondrial gene encoding apocytochrome b was identified by heterologous hybridisation with the corresponding yeast and bovine probes (Dawson et al 1984). The maize COB gene has a protein coding sequence of 1164 bp and, unlike the homologous gene in S. cerevisiae which contains a strain-dependent number of introns [for example 'long' strains KL 14-4A and 777-3A COB contains five introns (Borst 1980) whereas the 'short' strain D273 10B contains only two and lacks the first three (Labouesse and Slonimski 1983)], the maize gene is not split.

Transcript analysis of the maize COB gene was undertaken for the following reasons:

1) To determine whether the putative COB is actively transcribed in the mitochondrion.

2) To identify the transcripts of the maize COB gene and to undertake some preliminary mapping experiments.

The M13mp8 clones containing defined restriction fragments of the maize COB region used in this study were kindly provided by

Dr. A.J. Dawson.

4.5.2 Transcript analysis of the maize COB gene

Three COB protein coding sequence specific probes were used to analyse the transcripts of the putative maize COB gene:

1) pZmEH680; a 680 bp EcoRI/HindIII clone of mt DNA in pBR328. The insert of this clone contains 680 bp of COB sequence completely internal to the protein coding sequence

2) M13 clone 979

3) M13 clone 640. Both M13 clones contain the same insert as pZmEH680. Clone 640 contains mt DNA inserted in the M13 vector mp8 and the sequence of the newly synthesised DNA strand from this template (when ^{32}P labelled by second strand synthesis) is identical to that of the COB mRNA. Hence the probe should only hybridise to transcripts from the 'non-coding' strand. Clone 979 contains the same DNA fragment in mp9. The insert is therefore in the opposite orientation and should hybridise to COB transcripts. The genomic positions of the DNA probes used in this study are shown in Fig. 4.26.

40µg of maize mt RNA was loaded into a single well 5cm wide, electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel and transferred to nitrocellulose. Parallel strips were cut from the blot and hybridised with the three probes pZmEH680, 979 and 640. The hybridisation pattern obtained is shown in Fig. 4.27.

The transcript patterns identified with nick-translated pZmEH680 and clone 979 labelled by second strand synthesis are identical. Both identify a major transcript of ca 2.25 K bases (presumed to be the mRNA), an additional strongly hybridising band of ca 4.2 K bases and numerous smaller RNA species. A longer exposure of the autoradiogram

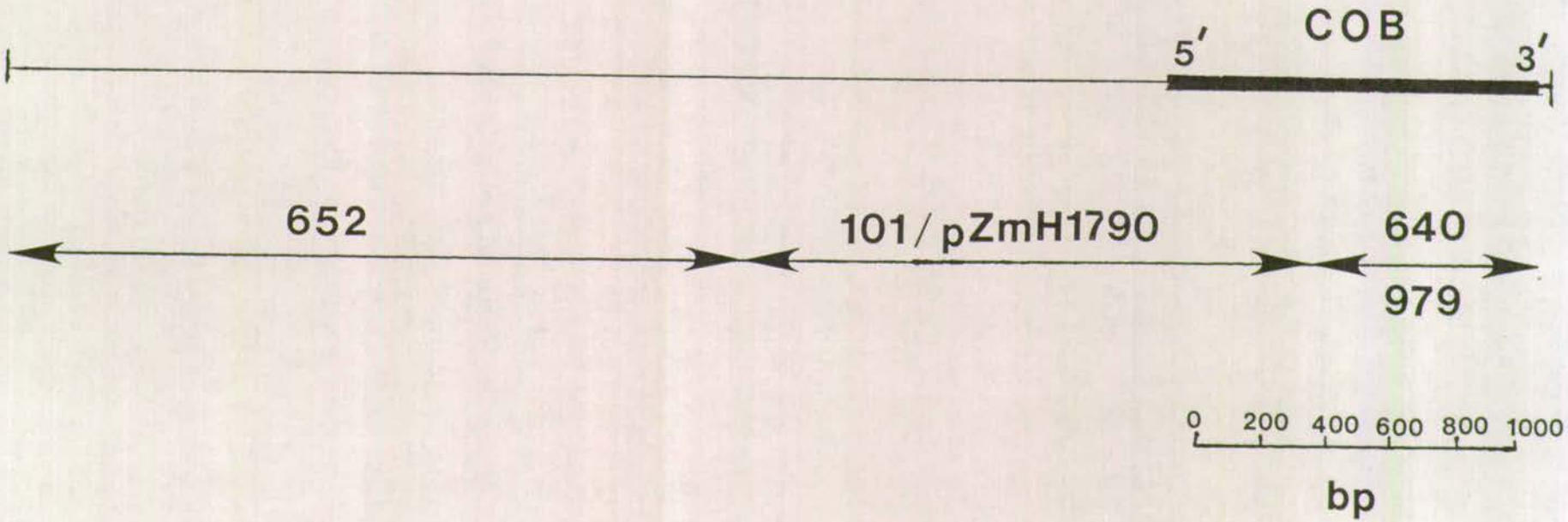


Fig 4.26 Origin of the probes used to characterise COB transcripts

Thickened bar represents the protein coding sequence.
Arrows delimit the inserts in the M13 clones used.

Fig. 4.27 Identification of transcripts of the maize COB gene

Maize mt RNA was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with clones 979 (track a), pZmEH680 (track b) and 640 (track c). RNA markers were run in parallel tracks, excised from the gel after electrophoresis and stained with ethidium bromide separately. RNA markers were cowpea chlorotic mottle virus (CCMV), E. coli, maize mt RNA, and tobacco mosaic virus (TMV).

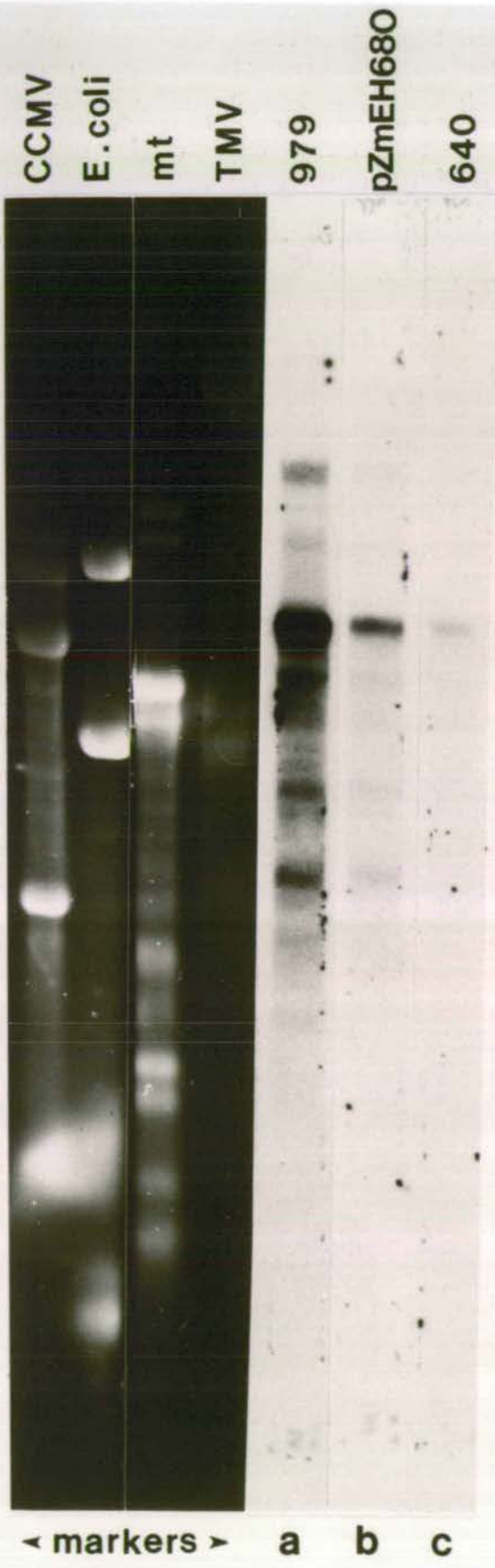


Fig. 4.27

reveals a very high molecular weight transcript homologous to clone 979. The molecular weight of this transcript can only be very approximately estimated due to the lack of suitable RNA markers in this size range. However, it may be greater than 9 Kb. The fact that both a pBR328 derived clone and an M13 clone gave an identical hybridisation pattern shows that:

1) vector homology makes no significant contribution to the pattern observed and 2) the majority of transcripts hybridising must come from the COB encoding strand. Clone 640 which had the same total and specific activity as 979 hybridised only weakly to the RNA blots, and the hybridising RNA species co-migrated with those detected with clone 979 and pZmEH680. It is thus very unlikely that these represent genuine transcripts from the opposite strand. As described in section 4.4.6.2, this apparent homology probably arises because the single stranded non-labelled template hybridised to the RNA is annealing with the M13 sequences in the ^{32}P labelled strand generated by second strand synthesis. The amount of hybridisation is lower presumably because a second order reaction is involved rather than a first order reaction as for directly complementary sequences. Thus any transcripts encoded by the opposite strand would be detected as more strongly hybridising bands on the fainter background. Since these are not seen, it is very unlikely that the non-coding COB strand (which contains an open reading frame of 186 bp (A.J. Dawson pers. commun.) is transcribed in the tissue studied.

4.5.3 Transcript analysis of COB transcripts using an upstream probe

Preliminary investigations into the map positions of the various COB transcripts were carried out using the M13mp9 clone, 652. This

This clone contains a 2.5 Kb EcoRI/HindIII insert, the 3' end of which lies 1370 bp 5' from the COB AUG initiation codon (Fig. 4.26).

Mt RNA was electrophoresed through 1.3% (w/v) agarose/7% (v/v) formaldehyde gel and blotted to nitrocellulose. Parallel strips were excised and probed with clones 652, 979 and 101 (an M13 clone containing the same insert of pZmH1790). The transcripts identified by each probe are shown in Fig. 4.28.

Clones 101 and 979 identify an identical pattern of transcription to that shown in Fig. 4.27. Hybridisation to the upstream clone 652 reveals hybridisation to a number of high molecular transcripts which also hybridise to pZmEH680 but does not identify the 2.25 Kb transcript presumed to be the mRNA.

4.5.4 Discussion

4.5.4.1 Summary of results

Transcript analysis of the maize COB gene has shown

1) mitochondrial RNA hybridises to COB protein-sequence specific clones showing the gene is probably transcribed in four day old etiolated maize coleoptiles

2) the COB protein coding sequence-specific probes identify a complex pattern of transcripts ranging in size from ca 1.0 to up to ca 9 Kb.

3) the major transcript, presumed to be the mRNA is much larger (ca 2.25 Kb) than is required to specify the COB polypeptide (1164 bases). Thus the presumed mature mRNA carries ca 1.4 Kb of additional sequence.

4) the DNA strand complementary to the COB coding strand is probably not transcribed in the tissue studied

Fig. 4.28 Transcript analysis of COB transcripts using an upstream clone

Maize mt RNA (10µg) was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with M13 probes from the maize COB region (see Fig. 4.26 for origin of these probes)

Lane a) clone 979
b) clone 101
c) clone 652

Fig.4.28

5) preliminary mapping studies have shown that a probe whose 3' end is ca 1.37 Kb from the AUG initiation codon identifies the high molecular weight transcripts which hybridised to the internal COB probes but does not identify the 2.25 Kb transcript presumed to represent the mature mRNA.

4.5.4.2 Multiple COB transcripts

The pattern of transcripts hybridising to the COB protein coding sequence specific probes is very complex, and is similar in its complexity to the transcripts of the maize mt COII gene. As for COII (section 4.3.6.2), several models can be proposed to account for multiple COB transcripts.

1) Presence of an intron

The multiple transcripts of the maize COB gene resemble those of the split COII gene in their complexity. Whilst it has been shown in wheat mitochondria that the presence of an intron in COII does not necessarily result in a complex pattern of transcription (Bonen et al 1984) it is possible that in maize mitochondria it does.

An obvious conclusion therefore is that the maize COB gene contains an intron which has not been detected by sequence analysis. Several features of the COB DNA sequence suggest that it may contain an intron of 806 bp at its 5' end (Dawson et al 1984, A.J. Dawson pers. commun.). These are as follows:

a) The initiation codon has a non-homologous position when compared to other aligned COB sequences.

b) The DNA sequence homology of the maize gene is low (ca 20%) compared to bovine and yeast COB sequences in the first 57 bases.

Otherwise homology is ca 50%.

c) The presence of a 36 nucleotide sequence 5' from the AUG initiation codon which could encode a short peptide with an initiator methionine almost coincident with that of other COB sequences and with good homology to the first amino acid residues in other species. The splicepoint which maximises homology would generate an intron of 806 bp.

The existence of this intron could be demonstrated by hetero-duplex analysis of DNA:RNA hybrids, possibly using the putative 2.25 Kb mRNA eluted from a gel. However, the presence of an intron at the 5' end of the maize COB is unlikely since recent sequence analysis of the Oenothera COB gene has revealed high homology at the amino terminus with the maize COB gene (A. Brennicke pers. commun.).

2) As suggested to explain the complexity of COII transcripts (section 4.3.6.2) the COB gene may be transcribed between multiple promoters and/or terminators. Alternatively a single large RNA molecule (perhaps the ca 9 Kb species) could be transcribed between a single promoter and terminator and then processed via a series of stable intermediates of lower molecular weight to yield the mature message.

Identification of the COB primary transcript(s) by the 'capping' assay would help resolve these possibilities.

4.5.4.3 Large size of the COB transcripts

The additional coding capacity of the large presumed mature mRNA could be occupied by sequences encoding other genes, ie the COB mRNA could be part of a polycistronic message. However, no long

open reading frames are found in the 1.5 Kb of sequence determined 5' to COB (A.J. Dawson pers. commun.). The presence of downstream open reading frames cannot be excluded, because of the lack of sequence data. However, stable polycistronic mRNAs have not yet been found in maize mitochondria. It has been shown (sections 4.3 and 4.4) that both COII and COI have transcripts far longer than required to specify the corresponding polypeptides, yet the 5' and 3' flanking regions of COI (P.G. Isaac pers. commun.) do not contain any long open reading frames.

The pattern of transcripts hybridising to the upstream clone 652 is striking in that it identifies the higher molecular weight but not the 2.25 Kb 'mature' mRNA. There are four possible interpretations of this result:

1) The maize COB gene is transcribed from multiple promoters. Thus the 9.0 Kb would represent the transcript originating from the most distal promoter. Clone 652 does not hybridise to the 2.25 Kb RNA because this transcript is initiated downstream of that region of the genome.

2) The maize COB gene is transcribed between a single promoter and terminator and processed via a series of intermediates of lower molecular weight. Clone 652 does not hybridise to the 2.25 Kb RNA because it is processed downstream of that site. Whilst these two models cannot be distinguished with the available data, any transcription initiation and/or processing sites could be more accurately mapped using a series of DNA probes specific for short regions of the COB genome.

3) The COB region is repeated in the maize mt genome. Thus for example the 2.25 Kb transcript could result from one copy of the

COB gene whilst the two higher molecular weight transcripts result from other copies of COB and the position of the promoter has changed in these copies eg by introduction of a stronger competing promoter. There is very preliminary evidence that COB sequences are repeated elsewhere on the maize mt genome since COB specific probes hybridise weakly to a number of EcoRI fragments even though the COB gene does not contain an EcoRI site (A.J. Dawson pers. commun.).

4) The highest molecular weight RNAs are transcripts of a gene upstream from COB whose 3' tail extends into the 5' end of the COB gene. This would explain the intensity with which clone 652 hybridises to the high molecular transcripts, while probe 979 hybridises only weakly. Thus the HMW transcripts could stop within clone 979 and hybridise only faintly to this clone because they have only a limited extent of homology. S1 nuclease mapping of the 3' ends of COB transcripts would reveal if this is so.

These four models are depicted on Fig. 4.29.

4.6 TRANSCRIPTS OF THE F_1 ATPase α SUBUNIT GENE

4.6.1 The F_1 ATPase α subunit gene

The discovery that the F_1 ATPase α subunit gene is encoded in the mitochondrial genome of all plants studied was surprising, since in all other organisms it is encoded in the nuclear DNA (Hack and Leaver 1983, Boutry et al 1983). DNA sequences homologous to the E. coli F_1 ATPase α subunit gene have been identified in maize mt DNA, but sequence analysis is incomplete. It is not known whether the F_1 ATPase α subunit gene in maize is split by intron(s).

Fig. 4.29 Diagrammatic representation of the various models proposed to explain the pattern of transcripts hybridising to COB region DNA probes

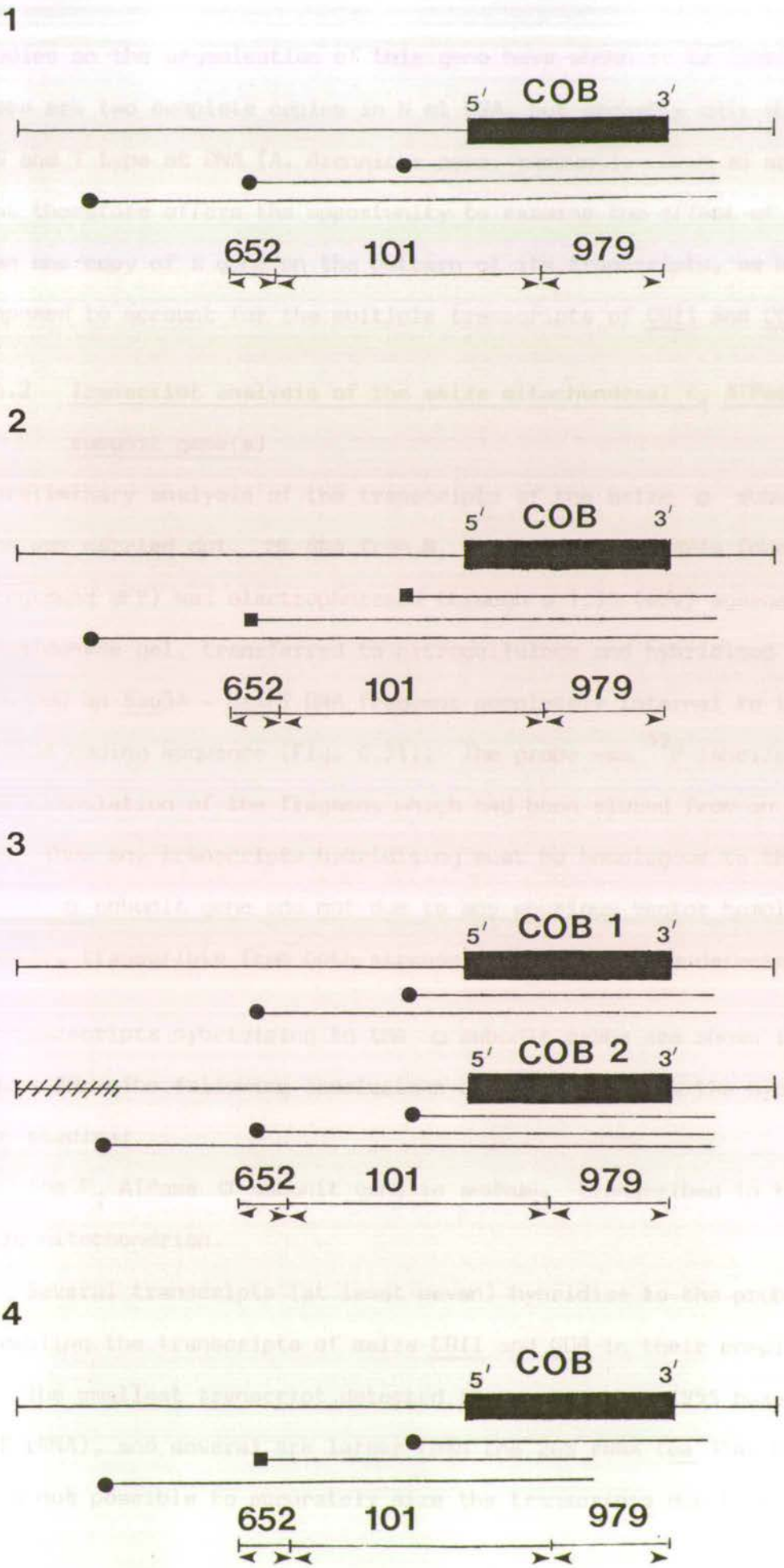
The pattern and relative stoichiometries of transcripts hybridising to probes 101, 652 and 979 could be explained by one of the four models depicted (diagrams not to scale).

Thickened bar represents protein coding sequence
● represents transcript originating from a promoter
■ represents a processed transcript

1) Multiple promoters
The multiple COB transcripts arise from initiation at multiple promoters
2) Single primary transcript and multiple processing sites
The largest transcript represents the primary transcript which is processed (trimmed) at many points to generate a series of lower molecular weight RNAs
3) Multiple copies of the COB gene
Copy No. 1 is transcribed from a single promoter. A DNA rearrangement places a second promoter 5' to the COB gene which is actively involved in transcription of COB copy No. 2.
4) The largest transcripts stop within COB protein coding sequence
The largest transcript stops within clone 979, thus explaining their weak hybridisation with this clone compared to the lower molecular weight transcripts. Clone 101 hybridises more strongly to the higher molecular weight transcript because it has higher sequence homology.

NB No experimental evidence for any of these models.
The numbers of the M13 clones used are above the bars showing their position.

Fig 4.2


1
5′ COB 3′
652 101 979
2
5′ COB 3′
652 101 979
3
5′ COB 1 3′
5′ COB 2 3′
652 101 979
4
5′ COB 3′
652 101 979

Studies on the organisation of this gene have shown it is likely that there are two complete copies in N mt DNA, but probably only one in C S and T type mt DNA (A. Brennicke pers. commun.). Such an arrangement therefore offers the opportunity to examine the effect of more than one copy of a gene on the pattern of its transcripts, as has been proposed to account for the multiple transcripts of COII and COB.

4.6.2 Transcript analysis of the maize mitochondrial F_1 ATPase α subunit gene(s)

A preliminary analysis of the transcripts of the maize α subunit gene was carried out. Mt RNA from N, C and S mitochondria (nuclear background WF9) was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with a ca 300 bp Sau3A - BamHI DNA fragment completely internal to the protein coding sequence (Fig. 4.31). The probe was ^{32}P labelled by nick translation of the fragment which had been eluted from an agarose gel. Thus any transcripts hybridising must be homologous to the F_1 ATPase α subunit gene and not due to any spurious vector homology. However, transcripts from both strands (if any) will be detected.

The transcripts hybridising to the α subunit probe are shown in Fig. 4.30. The following conclusions can be drawn from the hybridisation studies:

1) The F_1 ATPase α subunit gene is probably transcribed in the maize mitochondrion.

2) Several transcripts (at least seven) hybridise to the probe, resembling the transcripts of maize COII and COB in their complexity.

3) The smallest transcript detected is greater than 1955 bases (18S rRNA), and several are larger than the 26S rRNA (ca 3546 bases). It is not possible to accurately size the transcripts due to a lack

Fig. 4.30 Transcripts of the F_1 ATPase α subunit gene in maize mitochondria

10µg each of mt RNA from WF9 N, C and S type mitochondria was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with a nick translated DNA probe specific for the protein coding sequence of F_1 ATPase α subunit gene.

Fig. 4.31 Origin of the probe used to characterise F_1 ATPase α subunit transcripts

Thickened bar represents protein coding sequence.
Arrows delimit the length of the DNA probe used.

S = Sau3A site
B = BamH1 site

Fig. 4.30

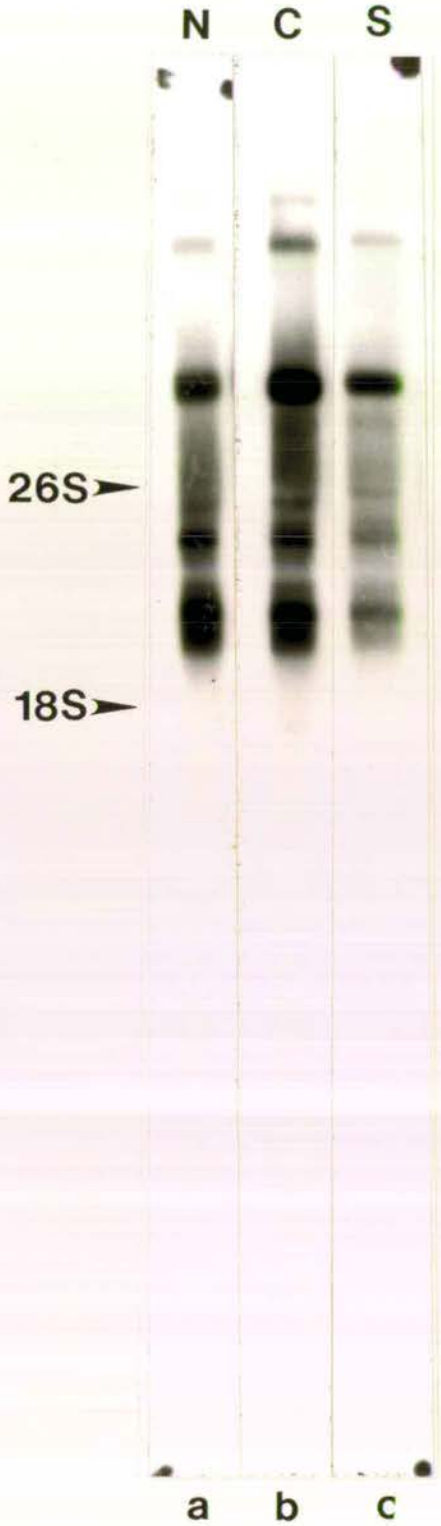
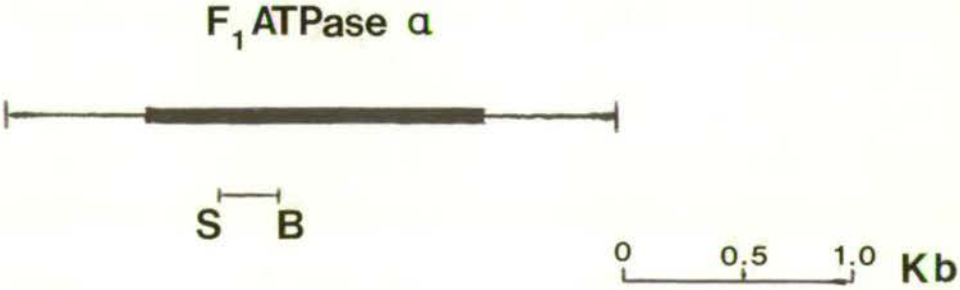


Fig. 4.31



of suitable size markers on the gel.

4) F_1 ATPase α subunit transcripts appear to be identical in N C and S type mitochondria. Since N mt DNA contains two copies of the gene, whereas C and S contain only one, then in N, either one of the copies is not transcribed (and therefore represents a pseudogene or a copy that is not expressed in four-day-old dark grown seedlings), or both are transcribed from identical promoters(s) and/or are processed at the same site(s). Further analysis of the sequences around both copies in N together with an accurate map of the transcripts is required before predictions about the effects of multiple gene copies on expression can be made.

4.7 TRANSCRIPTION OF OTHER GENES AND GENERAL FEATURES OF MAIZE MITOCHONDRIAL TRANSCRIPTS

4.7.1 Transcript analysis of the maize episomal DNAs S1 and S2

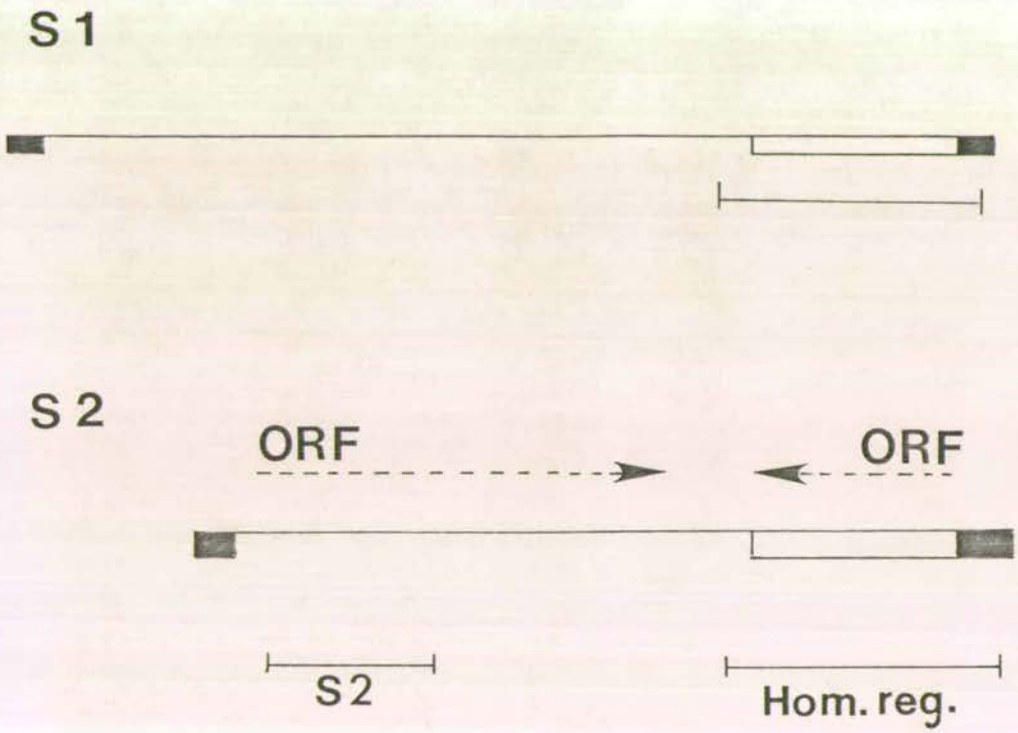
General features of the episomal DNAs S1 and S2 have been discussed in Chapter 1 (section 1.8.2.2). Transcript analysis of these episomal DNAs was undertaken with the following objectives:

1) To determine whether the episomal DNAs are transcribed both in the free form and integrated form and

2) To partially characterise these transcripts.

The episomal DNAs S1 and S2 have been completely sequenced (Levings and Sederoff 1983, C.S. Levings pers. commun.) and both shown to contain open reading frames (Fig. 4.32). As yet it is not known whether these open reading frames are transcribed. Recent studies on the presence of integrated forms of S1 and S2 in the maize mitochondrial genome suggest that characterisation of the episomal DNA transcripts may be more complex than previously thought. The salient

Fig. 4.32 Sketch map of episomal DNAs S1 and S2 to show origin of the probes used in transcript mapping, and location of the open reading frames.



Solid thickened bar represents the terminal inverted repeat
Open thickened bar represents the S1/S2 homologous region

features are as follows:

1) S1 and S2 only exist 'free' (ie non-integrated with 'mainband' mt DNA) in cms S and Vg cytoplasms

2) S1 and S2 sequences are integrated into 'mainband' mt DNA in

a) N type cytoplasm; integrated copies lack the 1500 bp homologous region (Thompson et al 1980, McNay et al 1983)

b) S and Vg type cytoplasms. Complete copies of S1 and S2 are integrated into 'mainband' mt DNA (Schardl et al 1984a). The 208 bp terminal inverted repeats (Levings and Sederoff 1983) are found integrated at many points in 'mainband' S type mt DNA (Schardl et al 1984a, see also section 4.4.6.6)

3) S type mt DNA appears to exist in a linear (as opposed to circular) configuration. This is thought to be caused by recombination of the integrated inverted repeat sequences (2b above) with free S1 and S2. The linear molecules have S1 and S2 terminally linked at their ends. S1 and S2 can be attached in either orientation (Schardl et al 1984a)

4) Cytoplasmic reversion to fertility is correlated with a loss of free S1 and S2 and a reversion of linear mitochondrial DNA to the circular form. Integrated copies of S2 are altered by a displacement of the 208 bp terminal inverted repeat at the unique sequence end of S2. Thus any promoters in this region would be inactivated. No major alterations of integrated S1 sequences in cytoplasmic revertants have been detected (Schardl et al 1984b)

5) Sequences homologous to S1 but not S2 have been detected in nuclear DNA from N, SRf, cms S and nuclear revertant plants (Kemble et al 1983) although the possibility that this in fact results from the integration psbA chloroplast gene in the nuclear genome (part of which is carried on S1 (section 1.9 and S. Levings pers. commun.) has

not been investigated.

It is apparent from these data that the episomal DNAs S1 and S2 can exist in many 'genomic locations' in the cell ie as free episomal DNAs, integrated into the mainband mt DNA, integrated into mainband mt DNA but in an altered form, and integrated into the nuclear DNA (S1 only). However the only cytoplasmic genotype where S1 and S2 are present in complete copies (both free and integrated) is in cms S and cms S (nuclear revertant). In all other cytoplasms, free S1 and S2 are absent, and integrated forms are altered in some way.

For these reasons it is clear that an analysis of the S1 and S2 transcripts (if any) is a complex undertaking since integrated copies and partial copies may be transcribed from different or additional promoters from free S1 and S2. Nevertheless the preliminary data presented in this section suggest this may be a fruitful area for future research.

Mt RNA was prepared from the following maize lines

N cytoplasm

Vg sterile No. 801-3

Cytoplasmic revertant No. 793

Cytoplasmic revertant No. 799

Cytoplasmic revertant No. 801-3

Nuclear revertant RfH

DNA was removed from the mt RNA preparations by treatment with DNase

The DNase-treated mt RNA was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, blotted to nitrocellulose and hybridised with DNA probes specific for

a) S2

b) the 1500 bp homologous region.

The locations of these probes are shown in Fig. 4.32. Fig. 4.33 and Fig. 4.34 show the transcripts hybridising to these probes.

The hybridisation data is summarised in Fig. 4.35 and Fig. 4.36. The S2 specific probe hybridises strongly to two transcripts of ca 5.2 and 4.7 Kb in Vg cytoplasm and the nuclear revertant, and weakly to a high molecular weight transcript in N. A probe specific for the 1500 bp homologous region hybridises to a complex array of transcripts from all the cytoplasmic and nuclear genotypes.

4.7.1.1 Discussion

Transcript analysis of the episomal DNAs S1 and S2 has shown

1) S2 is transcribed in Vg sterile cytoplasm and nuclear revertant. Two transcripts of 4.7 and 5.2 Kb hybridise to a DNA probe specific for S2. A very large transcript in N mt RNA also hybridises.

2) A DNA probe specific for the 1500 bp homologous region shared by S1 and S2 hybridises to many transcripts which differ in the cytoplasmic genotypes analysed

a) One transcript of 2.5 Kb is common to N, Vg, cytoplasmic revertant and nuclear revertant. Since 'free' S1 and S2 are not found in cytoplasmic revertant or N type mitochondria, this transcript must originate from either mitochondrial or nuclear integrated copies. Hybridisation of a homologous region probe to N mt RNA is surprising since in N mt DNA the homologous region of S1 and S2 is absent in the integrated forms (McNay et al 1983).

b) At least three transcripts of 2.5, 3.1 and 3.4 Kb are common to Vg sterile and cytoplasmic revertant. These must be transcribed from integrated S1/S2 in the cytoplasmic revertant since free S1 and S2 are absent in cytoplasmic revertants.

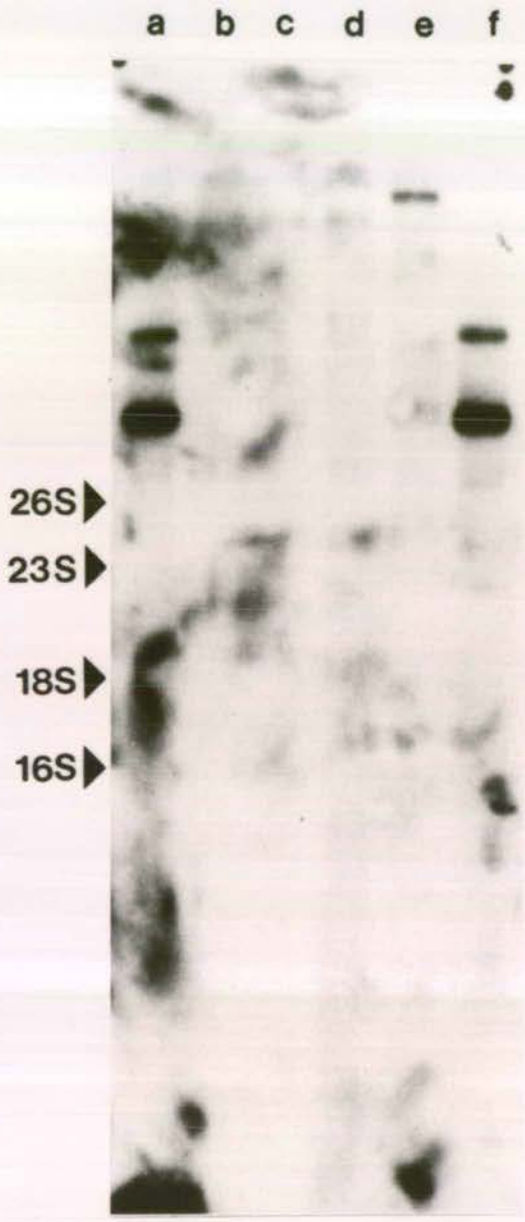


Fig 4.33 Transcripts hybridising to a DNA probe specific for episomal DNA S2.

Mt RNA was isolated from N and Vg sterile lines outlined below, electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with an S2 specific DNA probe (Fig 4.32).

Track a) Nuclear revertant Rf-H
b) Vg cytoplasmic revertant No.801-3
c) Vg cytoplasmic revertant No.799
d) Vg cytoplasmic revertant No.793
e) N
f) Vg sterile No/801-3

10 µg of RNA was loaded in each track.

Fig. 4.34 Transcripts hybridising to a DNA probe specific for the 1500 bp homologous region of episomal DNAs S1 and S2

Mt RNA was isolated from N and Vg sterile lines outlined below, electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose and hybridised with a DNA probe specific for the 1500 bp homologous region of S1 and S2 (Fig. 4.32).

Track a) Nuclear revertant RfH
Track b) cytoplasmic revertant No. 793
Track c) cytoplasmic revertant No. 799
Track d) cytoplasmic revertant No. 801-3
Track e) Vg sterile No. 801-3
Track f) N

10µg of RNA was loaded in each track.
Markers (*E. coli* RNA and maize mt RNA) were electrophoresed in parallel tracks and excised and stained separately.

Fig. 4.34

Fig. 4.35 Transcripts hybridising to an S2 specific probe: summary of data

Cytoplasmic genotype	Free S1 and S2	Integrated S1 and S2	Transcripts Kb
N	-	+ (S1 and S2 altered)	+?
Vg sterile	+	+	5.2, 4.7
Cytoplasmic revertant	+	+ (S2 altered)	-
Nuclear revertant	+	+	-

Fig 4.36 Transcripts hybridising to 1500 bp homologous region probe: summary of data

Cytoplasmic genotype	Free S1 and S2	Integrated S1 and S2	Transcripts Kb
N	-	+ (S1 and S2 altered)	2.5
Vg sterile	+	+	1.1, 1.5, 2.5, 3.1, 3.4, 4.7
Cytoplasmic revertant	-	+ (S2 altered)	2.5, 3.1, 3.4, and others of > 4.7 Kb
Nuclear revertant	+	+	1.1, 1.5, 2.5, 3.1, 3.4, 4.7

It is clear from these data that before further predictions about the transcription of S1 and S2 and the effects of integration can be made, a complete transcription map of free S1 and S2 must be constructed. In the absence of an *in vitro* transcription system for analysing run-off transcripts, this could be achieved by hybridising a series of overlapping single stranded M13 clones of S1/S2 and hybridising them to identical blots of cms S mt RNA. Since S episomal DNAs are present in a five molar excess over mainband mt DNA, any contribution from integrated copies is likely to be minimal. A more ambitious approach would be to use a homologous *in vitro* transcription system. This would potentially be more useful since the primary transcripts of S1 and S2 alone could be studied.

4.7.2 Variation in mt RNA profiles between N, Cms C S and T lines of maize

One particular S type cytoplasm (LBN) carries two double stranded RNAs LBN1 and LBN2 of 2.9 and 0.75 Kbases respectively (Sisco *et al* 1984). These are related to two single stranded RNAs of the same size found in other S cytoplasms (Schuster *et al* 1983), and also share homology with each other. These RNAs do not show any homology to mainband mt DNA nor to the episomal DNAs S1 and S2, and their origin and function remains unclear.

Examination of the total RNA in N and cms C S and T mitochondria reveals two abundant RNAs (denoted by arrows in Fig. 4.37) which are not detected in mt RNA isolated from N C and T type mitochondria. Very approximately these RNAs are 0.8 and 3.0 Kbases, suggesting they may be homologous to the ds RNAs detected in LBN cytoplasm and single stranded RNAs in other S cytoplasms.

Fig 4.37 RNAs peculiar to S type mitochondria

RNA was isolated from WF9 N,C,S,and T type mitochondria, denatured with glyoxal and electrophoresed through a 1.25% (w/v) agarose gel. Arrows point to the two RNA species of ca. 3.0 and 0.8 Kb. which seem to be unique (or over represented in) S type mt RNA.

4.7.3 Maize mitochondrial transcripts do not contain detectable poly A^+ tails

To investigate whether maize mt RNAs are polyadenylated as in human mitochondria (Hirsch and Penman 1974) 10µg maize poly A^+ RNA (a kind gift from A. Baker) was electrophoresed through a 1.3% (w/v) agarose/ 7% (v/v) formaldehyde gel, blotted to nitrocellulose and hybridised with a COB specific probe. No hybridising RNAs were detected. Since maize mt mRNAs can just be detected from 10µg total cytoplasmic RNA, it is unlikely that they are polyadenylated since if they were, they would be enriched by an oligo dT column. Yeast mt RNAs have similarly been shown not to be polyadenylated since a) yeast mt RNAs are not retained on an oligo dT column (Moorman et al (1978) and b) the 3' ends of many transcripts have been mapped and shown to co-align exactly with the mt DNA (see for example Coruzzi et al 1981). Alternatively it is possible that maize mt RNAs have short poly A^+ tails that are not retained by an oligo dT column. If so, an alternative approach to showing whether maize mt RNAs are polyadenylated would be to use synthetic oligo dT primers and synthesise a second strand using purified RNAs as templates. Comparison of the second strand sequence with that of the DNA would show whether maize mt RNAs do contain short poly A^+ tails.

4.8 OVERALL SUMMARY, DISCUSSION AND FUTURE PROSPECTS

The characterisation of the transcripts of the maize mitochondrial genes identified to date has made a significant major contribution to our overall understanding of the plant mitochondrion and has raised many important questions which are worthy of future investigation.

The characteristics of maize mt transcripts analysed in this chapter are summarised below:

1) The mt genes encoding cytochrome oxidase subunits I and II, apocytochrome b and F_1 ATPase α subunit are transcribed in the maize mitochondrion

2) The patterns of transcripts hybridising to COB, COII and α ATPase probes are very complex and resemble transcripts of the yeast mt genes which contain introns for example COB and COI. However it is likely that of the maize mt genes analysed, only COII contains an intron, and thus a complex pattern of transcripts does not necessarily result from excision of an intron. This is reinforced by the finding that the wheat COII gene is seemingly transcribed as a single precursor RNA which is then spliced to generate a single mRNA and a stable intron encoded RNA (Bonen et al 1984). It is striking that the wheat and maize COII genes are highly conserved ($>$ 99% in the protein coding sequence, ca 99% in the intron (excepting the intron insert) suggesting there are tight constraints on exon and intron coding sequence, and yet their pattern of transcription is so dissimilar. The possibility that the open reading frame 5' of the maize COII gene encodes part of a maturase has been discussed.

A model has been proposed to account for these complex patterns of transcripts ie they arise either from processing of a single large primary transcript or from transcription from many promoters. Examples of both of these are found in yeast mitochondria. Tandem promoters are found upstream of the ATPase 9 gene (Edwards et al 1983), and yeast COII gene probes hybridise to a complex array of transcripts, even though the gene does not contain an intron (Corruzzi et al 1981).

The discovery of COII homologous sequences in the maize mt genome may complicate the pattern of transcripts observed.

3) The maize COI gene hybridises to two major RNA species. Nuclease S1 mapping of the 5' ends of these transcripts has shown that the COI transcripts may be more heterogeneous than revealed by Northern blotting, and there may be up to five RNAs transcribed from this gene. Primer extension would show whether these observations are real phenomena.

The two major COI transcripts start at ca 152 and 57 bases 5' from the AUG initiation codon. The sequence around the start of the higher molecular weight transcript shows 8/9 homology to the nonanucleotide sequence thought to form part of a fungal mitochondrial promoter (Osinga and Tabak 1982, Osinga et al 1982, Tabak et al 1983) and 13/15 homology to the sequence around the 5' end of the transcript of a Beta vulgaris mt episomal DNA (Munk Hansen and Marcker 1984). The sequence at the 5' end of the COI lower molecular weight transcript shows 6/12 homology to a dodecamer sequence thought to be involved in processing (as opposed to splicing) of yeast mt RNAs (Osinga et al 1984). Since the primary COI transcript has not yet been identified it is not possible to conclude whether the two COI transcripts result from separate transcription initiation events, or whether the lower molecular weight COI RNA results from a processing event.

Heterogeneity such as that of the maize COI RNAs is not unprecedented in plant mitochondria (Spencer et al 1984, Schnare and Gray 1982). Post transcriptional processing and trimming may be a general feature of plant mitochondrial RNAs, and may represent another point at which gene expression is controlled. In yeast the 15S rRNA is similarly

'ragged' suggesting that there is imprecise processing of this RNA (Christianson et al 1981).

4) The majority of transcripts hybridising to each of the gene specific probes are far longer than required to encode the corresponding polypeptide. This is also a feature of many fungal mt RNAs [for example COI in Neurospora crassa (6Kb, De Jonge et al 1983), COB in S. cerevisiae (ca 2 Kb, Borst and Grivell 1981a), ATPase 9 in S. cerevisiae (820 nucleotides-protein coding sequence is 225 bp, Hensgens et al 1979)]. The two major COI transcripts carry between 700 and 600 bases of additional information, with approximately 550 - 450 bases as a 3' terminal extension and leaders of ca 152 and 57 bases. The map positions of the transcripts of COII, COB and ATPase have not been determined.

The length of the maize mt RNAs suggests they could form polycistronic transcripts. However, the available sequence data suggests that this is not so, since no significant open reading frames have been found upstream or downstream of the genes analysed.

5) The transcripts of COI are not altered in N C S and T mt RNA. Similarly no differences were detected in N C and S mt RNA for F_1 ATPase α subunit transcripts.

COII transcripts show considerable differences from N in C S and T cytoplasms. The cause of these differences is not known. An understanding of the basis of these differences requires a) more detailed analysis of the DNA rearrangements around the COII gene in C, S and T mt DNA and b) identification of the COII primary transcript(s).

6) The episomal DNAs S1 and S2 are transcribed in a complex manner, and both integrated and free forms are apparently transcribed. These patterns of transcription are likely to be complicated by the altered integrated forms of S1 and S2.

7) In common with yeast mt RNAs, maize mt RNAs do not appear to have detectable poly A^+ tails.

8) The intron of COII is apparently excised from COII precursor RNAs and remains in the mitochondrion as a stable molecule. The physical form of this stable RNA is not known.

4.8.2 Future prospects

The work described in this chapter has initiated a large and interesting area of plant molecular biology, and has posed many more questions than it has answered. An understanding of the expression of the plant mitochondrial genome will initially require more extensive mapping of transcripts to the genome. However, many more biochemical analyses will be required before predictions about the effects of DNA rearrangements on mt gene expression and the way in which higher plant mt gene expression is controlled can be made. Two such approaches which could be used in the near future are outlined below.

4.8.2.1 Identification of primary transcripts

In order to identify plant mt promoters, and any polycistronic mRNAs, the primary transcripts of the mt genes must be identified since the transcript termini identified by S1 nuclease protection will not necessarily identify the points at which transcription is initiated. Only when these have been characterised can predictions about any processing pathways be made. There are two approaches to identifying

primary transcripts:

1) All the primary transcripts of the maize mt genome could be identified by the guanylyl transferase assay. Thus total mt RNA would be labelled using guanylyl transferase and Y GTP. Electrophoresis of the labelled RNA followed by autoradiography would theoretically identify the number of primary transcripts. Partial hydrolysis of the RNA followed by hybridisation to a cosmid or plasmid library of mt DNA would identify the regions of the mt genome which contain promoters. This approach has been used to identify the primary transcripts of the yeast 21S and 15S rRNAs (Levens et al 1981, Christianson et al 1982). However, in yeast, many primary transcripts are so rapidly processed that they escape detection by the guanylyl transferase assay, and thus they must be enriched in some way (Christianson et al 1983).

2) An enrichment of primary transcript(s) for a particular gene could be achieved by hybrid release. This involves coupling a gene specific DNA probe to a solid support eg nitrocellulose, hybridising total mt RNA to the DNA, and then eluting the bound transcripts. The primary transcript(s) of that gene could then be labelled using the guanylyl transferase assay, purified by gel elctrophoresis and sequenced. Comparison of the RNA sequence with the DNA sequence upstream of the gene would identify the point of initiation. A potential disadvantage of this approach is that if as in yeast mitochondria, maize mitochondrial genes are co-transcribed from a common promoter, and if processing of transcripts occurs while they are still part of the transcription complex, then the primary transcript will not be identified.

4.8.2.2 In vitro transcription

The development of an homologous in vitro transcription system using

purified mt RNA polymerase would greatly enhance our understanding of the expression of the plant mt genome, and would overcome some of the potential problems in the identification of primary transcripts by the 'capping' assay. Thus mt DNA fragments could serve as templates. Such a system has been shown to initiate transcription with fidelity on yeast mt DNA templates (Edwards et al 1982). In addition to a) identification of regions of the mt genome giving rise to primary transcripts and b) identification of promoter sequences, the development of an in vitro transcription system has a number of uses relevant to the study of plant mitochondrial DNA. Thus for example it could be used to analyse a) the transcription of the episomal DNAs S1 and S2 in the absence of integrated copies, b) the transcription of repeated gene sequences for example the repeated COII, COI and α ATPase sequence(s) and c) the effect of DNA rearrangements on transcription of genes for example COII in T mt DNA.

4.8.2.3 Identification of genes

An in vitro transcription system could also aid in the identification of new mitochondrial genes. As discussed in Chapter 3, the plant mt genome is likely to contain genes not found in the mitochondria of other organisms, and so the scope for identification of genes by heterologous hybridisation is limited. Identification of transcriptionally active regions of the mt genome (either by in vitro transcription or by hybridisation ^{32}P end labelled RNA to mt DNA) followed by sequencing of the mt DNA could reveal the presence of new genes which could not be identified by any other means in the absence of a cell free in vitro translation system (Chapter 3). This is a circuitous way of identifying plant mt genes, since as has been shown in this chapter, many mt transcripts contain additional information with no obvious coding function

and large stretches of DNA would need to be sequenced to identify new genes. In order to determine whether the open reading frames identified in this way are actually expressed as polypeptides it would be necessary to prepare antibodies either by a) synthesis of large amounts of the polypeptide in an expression vector where the DNA of interest is fused with E. coli promoters and ribosome binding sites (Maniatis et al 1982) followed by antibody production or b) construction of synthetic polypeptides (whose composition is dictated by the DNA sequence of interest) which are used to raise antibodies.

4.8.3 Conclusions

In conclusion it is important to stress that the identification and characterisation of the transcripts of the maize mitochondrial genome has and will make only a small contribution to an overall understanding of the biogenesis of this organelle.

It is clear from the work presented in this chapter that there are potentially many points at which mt gene expression may be controlled. Thus the level of a particular mRNA could be modulated by the availability of RNA polymerases (nuclear and/or mitochondrially encoded), intron splicing enzymes (nuclear and/or mitochondrially encoded), processing (opposed to splicing) enzymes, the rate of splicing/processing, the stability of processed intermediates and mature mRNAs. The translation of these mature mRNAs could in turn be modulated by the availability of nuclear encoded ribosomal proteins.

More complete analysis of the maize mt transcripts will allow us to make predictions about which points these control mechanisms might operate. Moreover as will be discussed in Chapter 6, since the vast majority of plant mt proteins are encoded in the nucleus (Leaver and Gray 1982) and

transported into the mitochondrion, it is the identification of these genes (particularly those forming mosaic enzyme complexes) together with studies on the co-ordinate expression of the two genomes which will throw most light on the biogenesis of a functional organelle.

CHAPTER 5 DEVELOPMENTAL REGULATION OF HIGHER PLANT MITOCHONDRIAL GENOME EXPRESSION

5.1 INTRODUCTION

This chapter describes a preliminary investigation into the control of mitochondrial gene expression during mitochondrial biogenesis in higher plants. The mitochondrion has two main functions in the plant cell: firstly to generate energy in the form of ATP and secondly to provide the cell with carbon skeletons for a variety of synthetic reactions, for example acetyl coenzyme A for fatty acid synthesis. Clearly the energy and metabolite requirements of the plant cell will vary at different stages of the life cycle, most notably during germination and early seedling development. In contrast the mammalian mitochondrion exists in a relatively homeostatic environment and is not likely to be subjected to fluctuating demands such as are placed on the higher plant mitochondrion. The higher plant mitochondrion thus offers an ideal system for studying the control of mt biogenesis which, as outlined in Chapter 1, requires the co-ordinate synthesis of polypeptides encoded in both the nucleus and the mitochondrion.

In Saccharomyces cerevisiae the number and morphology of mitochondria is known to change under different environmental conditions (Stevens 1981). When yeast is grown in a glucose-rich medium, catabolite repression reduces the synthesis of certain respiratory enzymes and the substrate is partially fermented. The change from such 'glucose repressed' conditions to growth on low levels of glucose ('derepression') is accompanied by an increase in the number of mitochondria and a change in their morphology from highly convoluted to ellipsoid forms (Stevens 1981). Release from glucose repression is also accompanied

by an increase in the de novo synthesis of certain mitochondrially encoded proteins (1983).

In higher plants, mitochondrial biogenesis has been studied in two systems where a resumption of metabolic activity after a period of dormancy is accompanied by an increase in mitochondrial activity. These systems are discussed below.

Ageing of Jerusalem artichoke (Helianthus tuberosus) tuber tissue discs in water induces mitochondrial biogenesis and an increase in respiratory activity (Forde et al 1979). Incorporation of ^{35}S-methionine by isolated mitochondria into protein increases three-fold during a 26 hr period of ageing. This apparent increase, however, is largely attributable to a depletion of the endogenous methionine pool during ageing. However an analysis of the polypeptides synthesised by 'aged' and 'non aged' mitochondria reveals increases in the relative rates of synthesis of two mt translation products (M_r 17,000 and 34,500) and a decrease in the rate of synthesis of a polypeptide of M_r 19,000 during ageing.

Mitochondrial biogenesis has also been studied in rehydration and early development of Vicia faba (faba bean) seedlings (Leaver and Forde 1980, Dixon et al 1980). Mitochondria isolated from cotyledons in dry seeds possess low respiratory activity. During the first 24 hours of rehydration respiratory activity increases six-fold. E.M. analysis of the cotyledons shows that between 12 and 48 hours after rehydration, the mitochondria undergo distinct ultrastructural changes,

notably an increase in the number of cristae, the appearance of a more 'electron-dense' matrix and an increase in the number of ribosome-like particles. By 48 hours after rehydration the mitochondria appear to be fully developed.

During the first 24 hours of germination the ability of isolated V. faba mitochondria to incorporate ^{35}S-methionine into protein remains very low; thus the initial recovery of mitochondrial integrity and activity is achieved independently of significant mitochondrial protein synthesis. The ability of isolated mitochondria to synthesise protein increases ten-fold during the second day of rehydration and estimation of the matrix amino acid concentrations has shown that this increase is not simply due to a depletion of the mitochondrial methionine pool during germination. Analysis of the mt translation products synthesised at various stages of rehydration and early development shows marked qualitative changes in the spectrum of polypeptides made. Mitochondria isolated from cotyledons 96 hours after rehydration synthesise about twelve major polypeptides. Of these only six are major translation products in the mitochondria from dry or twelve hour imbibed cotyledons. The relative rate of synthesis of the remainder increases mainly between 12 and 48 hours after rehydration. Particularly marked is the relative increase in the rate of ^{35}S-methionine incorporation into subunit I of the cytochrome oxidase complex. This was shown by a) densitometric scans of autoradiograms of mitochondrially synthesised proteins and b) immunoprecipitation of ^{35}S labelled polypeptides with yeast antibodies (Dixon et al 1980, C.J. Leaver pers. commun.). This system thus offers an ideal opportunity for the study of gene expression during mitochondrial biogenesis, as discussed below.

5.2 ANALYSIS OF MT GENE EXPRESSION DURING GERMINATION AND EARLY DEVELOPMENT OF V. FABA

One approach to the study of mitochondrial biogenesis during germination and early development of V. faba is to investigate the transcripts of particular mitochondrial genes. Of particular interest is 1) the lag in the ability of mitochondria isolated from dry V. faba seeds to incorporate ^{35}S-methionine into protein and 2) the differential rates of synthesis of certain polypeptides (for example COI) during early development. Semi-quantitative analysis of the transcripts of mitochondrial genes at different stages of development should indicate the means by which mitochondrial gene expression is controlled. Thus, for example, the following factors could be relevant with respect to both the lag in mitochondrial protein synthesis and the differential synthesis of COI:

1) a requirement for the de novo transcription of mt genes. This in turn could be regulated by the availability of mt RNA polymerase. As yet a higher plant mt RNA polymerase has not been isolated or characterised, and it is not known whether it is nuclear or mitochondrially encoded. In yeast the levels of the nuclear encoded mt RNA polymerase differ at various stages of the life cycle (Levens et al 1982)

2) deficiencies in the translational machinery of mitochondria at early stages of development.

3) Translational control. This could result from a) the presence of a translation inhibitor/absence of a mt translation factor as in yeast where nuclear encoded proteins are known specifically to stimulate the translation of certain mt proteins (Muller et al 1984); or b) a requirement for the processing of precursor RNAs to generate a

translatable mRNA. The analysis of the transcripts of maize mt genes (Chapter 4) has suggested that (at least in maize) there may be many points at which mt gene expression may be controlled at the levels of transcription and post-transcriptional modification. Thus for example the quantities of particular members of the numerous RNA species detected from maize mitochondria with COB, COII and F_1 ATPase α subunit DNA probes could be generated both by changes in the rate of transcription initiation and by the rates of any processing steps.

The aim of the work presented in this chapter was to study gene expression during mitochondrial biogenesis with particular emphasis on germination and early development of V. faba seedlings. V. faba, rather than maize, was chosen as a developmental system because although studies have shown a similar lag in the ability of isolated maize mitochondria to synthesise protein, the spectrum of proteins synthesised does not appear to change qualitatively during germination (V.P. Jones unpublished data). Thus the maize developmental system does not offer the opportunity to investigate why certain mt polypeptides are differentially synthesised during germination.

Quantification of particular gene transcripts at various stages of germination should allow crude estimates of the relative expression of those genes to be made. Quantification can be achieved by in vitro translation or Northern blotting and these techniques have been widely used for the analysis of gene expression during chloroplast biogenesis [Silverthorne and Ellis (1980), Walden and Leaver (1981) and Smith and Ellis (1981)]. In vitro translation allows the amount of translatable mRNA for a particular polypeptide to be quantified, whereas transcript analysis by Northern blotting [either as a 'dot blot' (Thomas 1983)

or as a gel blot] does not discriminate between translatable and non-translatable RNAs. Northern blotting (as a gel blot) does have the advantage that precursor/product ratios of RNAs undergoing post-transcriptional processing can be examined. This could be important if non-translatable precursor RNAs (eg intron-containing or other pre-processed forms) are accumulated during, for example, seed formation then processed to translatable forms during germination.

The failure to develop an _in vitro_ translation system for the expression of higher plant mt mRNAs (Chapter 3, this thesis) means that translatable mRNAs could only be quantified by the isolation of mt polysomes, electrophoresis of the RNAs (mRNA being translated and rRNAs), transfer to nitrocellulose and hybridisation with gene specific probes.

In this study, the relative amounts of all the transcripts [ie precursor(s) and product(s)]of a particular mt gene were analysed by Northern blotting. However, as outlined below, quantitative or semi-quantitative estimation of the amount of an organelle-encoded mRNA at different stages of development presents particular problems. The approach used in this study was as follows:

1) total cellular RNA was isolated from _V. faba_ cotyledons at various stages of rehydration and germination as quantitatively as possible. [_V. faba_ seeds were germinated under conditions exactly the same as the seeds used for protein synthesis studies (Dixon _et al_ 1980, B. Forde pers. commun.)]

2) it was demonstrated that maize mt gene specific DNA probes will hybridise to 'dot blots' of _V. faba_ total cellular RNA since _V. faba_ mt gene DNA probes are not yet available. Hybridisation under

stringent conditions (Chapter 2, section 2.2.4.5.2) between maize mt gene DNA probes and V. faba mt transcripts is expected since the overall nucleotide homology between Pisum sativum (another legume) and maize COII is 92% (R. Wu pers. commun.).

3) attempts were then made to estimate the relative amounts of V. faba mt gene transcripts at various stages of rehydration and germination by gel electrophoresis of total cellular V. faba RNA, transfer to nitrocellulose and probing with ^{32}P labelled mt gene specific DNA probes from maize. Parallel experiments have shown (Chapter 4, Fig. 4.4) that mt gene transcripts can be detected in similar blots of maize total cellular RNA, although hybridisation signals are weak. [Maize mt RNAs have been estimated by semi-quantitative Northern blottings to comprise $< 1\%$ of total cellular RNA (V.P. Jones unpublished data).] RNA from different developmental stages could be loaded onto gels either a) on an 'equal cotyledon' basis ie each sample represents a fixed proportion of the total RNA isolated from each cotyledon or b) equal amounts of RNA from each developmental stage can be loaded. In fact if both types of series are used, they can form controls for each other.

5.3 ISOLATION OF V. FABA TOTAL CELLULAR RNA FROM COTYLEDONS AT VARIOUS STAGES OF REHYDRATION AND GERMINATION

V. faba seeds were germinated in darkness and at 0, 12, 24, 48 and 72 hours after initial rehydration. Seeds were harvested, seed coats removed and cotyledons rapidly frozen in liquid nitrogen. The frozen seeds were stored in sealed containers at -80°C until required.

Total cellular RNA was isolated from rehydrated seeds as described in Chapter 2 (section 2.2.2.2). Three volumes of RNA extraction

buffer per gram fresh weight of seed was added to each sample containing thirty cotyledons from 15 seeds, and the RNA extracted by grinding for precisely 15 seconds with a Polytron homogeniser at full speed. RNA was extracted from dry seed by grinding the seed to a fine powder in a seed grinder and adding three volumes of RNA extraction buffer per gram of seed. All samples were then extracted six times with an equal volume of phenol, precipitated with ethanol and dissolved in water.

Fig. 5.1 shows that the amount of total cellular RNA per cotyledon does not increase significantly over the developmental period studied. This contrasts sharply with that of other dicotyledonous species during germination eg Cucumis sativus where the amount of total cellular RNA increases ten - fold between days 0 and 5 and then decreases (Becker et al 1978). However cucumber is an epigeal seed ie the cotyledons form the first aerial leaves and hence undergo much differentiation. In contrast V. faba is a hypogeal seed and the cotyledons do not form the first aerial leaves but merely act as storage organs. Thus a great increase in the amount of cellular RNA per cotyledon would not be expected.

5.4 SEMI-QUANTITATIVE MEASUREMENT OF RNA BLOTTED TO NITROCELLULOSE

Semi-quantitative measurements of the amount of a particular gene transcript by Northern blotting has the following limitations:

1) The amount of RNA loaded onto a gel or dot blot is critical and must not saturate the blotting medium.

2) The amount of ^{32}P labelled DNA probe must not be limiting.

3) The amount of ^{32}P labelled DNA probe hybridising to the RNA sample can be measured in two ways: a) scintillation counting of

Fig. 5.1 Isolation of total cellular RNA from V. faba cotyledons

Total cellular RNA was isolated from V. faba cotyledons (cot) at various stages of rehydration and early development. The amount of RNA in each sample was estimated by measuring A_{260} of duplicate aliquots.

Fig. 5.2 Semi-quantitative measurement of COI transcripts in maize mt RNA

Varying amounts (0 - 20µg) of maize mt RNA were loaded onto nitrocellulose (after being adjusted to 2.5M NaCl) and hybridised with an M13 clone specific for the maize COI gene, ^{32}P labelled by second strand synthesis. The filter was washed and exposed to pre-flashed X-ray film for ten days at -80°C.

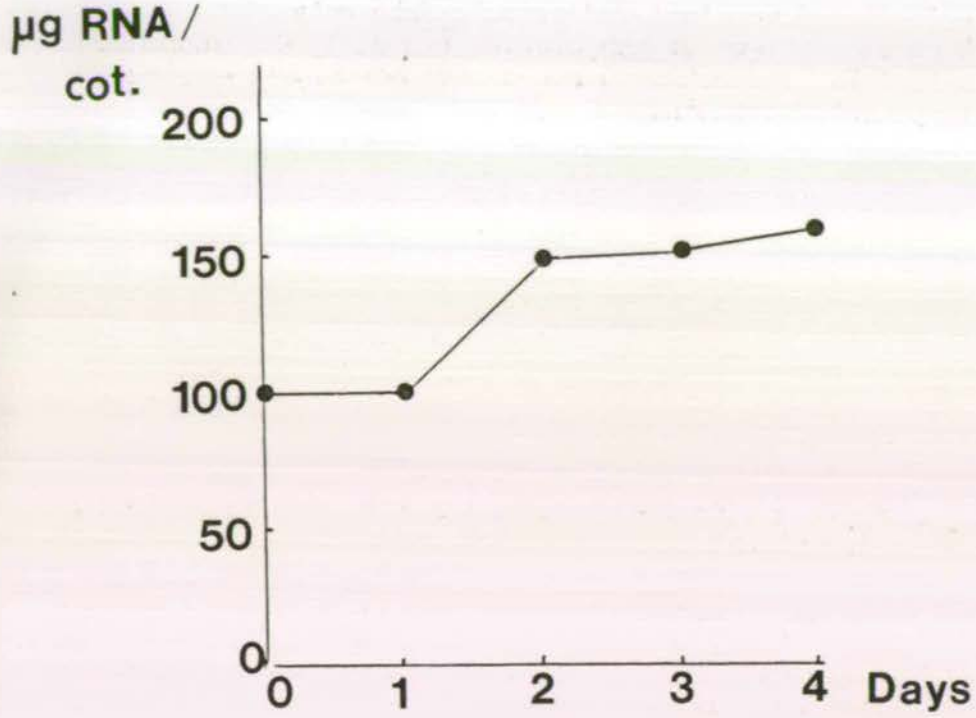


Fig. 5.1

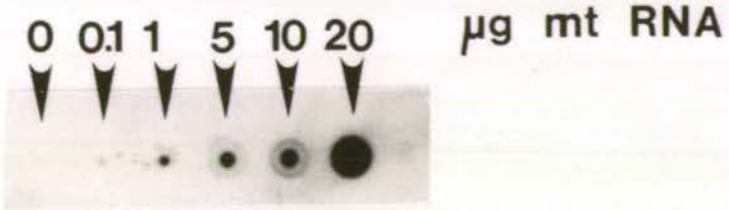


Fig. 5.2

radioactivity bound to the filter or b) use of X-ray film to detect the radioactivity. It is important in the latter case to demonstrate a linear response between the deposition of silver grains and the amount of isotope bound to the filter.

In view of the weak signals obtained when mt gene specific DNA probes are hybridised to total cellular maize RNA (Chapter 4, Fig. 4.4) semi-quantitative measurements of ^{32}P labelled DNA hybridised were made by X-ray fluorography rather than scintillation counting (Chapter 2, section 2.2.4.6). In order to check the linearity of response of the film, varying amounts of maize mt RNA (0 - 20 µg) were made 5M with respect to NaCl and loaded onto a nitrocellulose filter using a 'Hybridot' apparatus (Schliecher and Schull). The filter was hybridised with a ^{32}P labelled DNA probe specific for the protein coding region of the maize COI gene.

Fig. 5.2 shows that in this particular experiment, within the range 0µg - 20µg, a two-fold difference in the amount of a particular RNA species bound to nitrocellulose can be easily detected. However, due to variability in a number of factors such as film exposure time, a similar dilution series must be included as an internal control in exposures of developmental series.

5.5 HYBRIDISATION OF ^{32}P LABELLED MAIZE MT DNA PROBES TO V. FABA TOTAL CELLULAR RNA.

In order to determine whether ^{32}P labelled DNA probes specific for maize mt genes would hybridise to V. faba total cellular RNA, 5µg of V. faba total cellular RNA from various developmental stages was spotted onto nitrocellulose using the 'Hybridot' apparatus as described in section 5.4. In a parallel experiment 25µg of each

sample was electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel and transferred to nitrocellulose. Both blots were hybridised with a ^{32}P-labelled mt DNA probe specific for the maize COI gene (M13 clone labelled by second strand synthesis to ca 10^6 dpm/μg). Figs. 5.3 and 5.4 show the results of this hybridisation. The maize mt DNA probe hybridises to V. faba total cellular RNA on a dot blot but consistently fails to hybridise to a larger sample of RNA electrophoresed through an agarose gel and transferred to nitrocellulose. (It is important to note however that a control whereby non-specific hybridisation (eg to E. coli RNA) was estimated on the dot blot was not carried out.)

The consistent failure of Northern blots of V. faba total cellular RNA to hybridise to maize mt DNA probes could be the result of a number of factors:

1) A dot blot represents a high localised concentration of RNA whereas the transcripts of interest when electrophoresed through a gel are more diffuse. This will be particularly true if, as in maize, several transcripts with different electrophoretic mobilities arise from each gene.

2) Transfer of RNA from a gel to nitrocellulose is probably less efficient than direct application.

3) The nucleotide homology between dicotyledonous plant and monocotyledonous plant mt genes is too low to generate a strong hybridisation signal with total cellular V. faba RNA. Lowering the hybridisation stringency could overcome this problem (Bonner et al 1983, Casey and Davidson 1977). Alternatively it may be necessary to isolate mitochondrial RNA from V. faba cotyledons (rather than total cellular RNA) and examine any changes in the relative proportion of

Fig. 5.3 Hybridisation of V. faba total cellular RNA with maize COI gene specific DNA probe

Duplicate samples (5µg) of RNA isolated at various stages of rehydration and germination of V. faba cotyledons were loaded onto nitrocellulose and hybridised with an M13 clone specific for the maize COI gene, ^{32}P labelled by second strand synthesis. The filter was washed and exposed to pre-flashed X-ray film for fourteen days at -80°C.

Fig. 5.4 Analysis of V. faba COI transcripts by Northern blotting

Maize mt RNA (5µg) and V. faba total cellular RNA (25ug) were electrophoresed through a 1.3% (w/v) agarose/7% (v/v) formaldehyde gel, transferred to nitrocellulose, and hybridised with an M13 clone specific for the maize COI gene, ^{32}P labelled by second strand synthesis. The filter was then exposed to X-ray film at -80°C for fourteen days.

Note that under the conditions of hybridisation/washing, the maize DNA probe consistently fails to hybridise to V. faba total cellular RNA.

Fig. 5.3

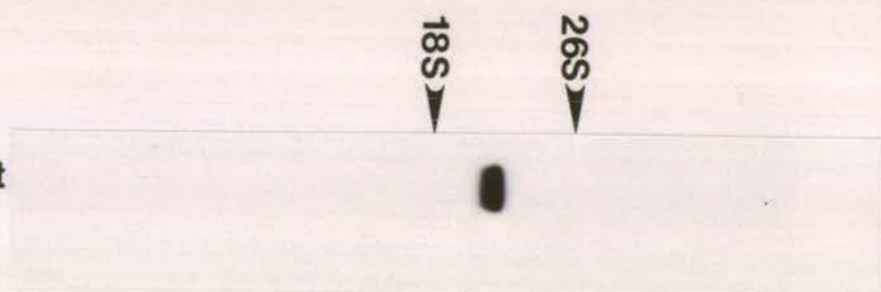


Fig. 5.4

particular gene transcripts within a fixed mass of RNA.

Fig. 5.3 shows equal amounts (5µg) of V. faba total cellular RNA isolated from cotyledons at various stages of germination hybridised with COI specific ^{32}P labelled DNA probe (in duplicate). Since a number of controls were missing from this experiment, it is not possible to draw any firm conclusions about the control of mt gene expression during germination and early development of V. faba seedlings. Thus for example 1) a dilution series was not included in this particular experiment; 2) non-specific hybridisation was not investigated and 3) the contribution of V. faba DNA to the hybridisation signal was not measured (since DNA was not removed from these samples). However in spite of these absent controls very preliminary results suggest that the COI transcript(s) are present at low levels in dry V. faba seed and that the level of these increases within 12 hours of rehydration ie preceding the increase in mt protein synthesis and the synthesis of COI. This experiment should be repeated with the controls listed above in order to draw firm conclusions on this data. Also, a dot blot does not show whether any change in the relative amount of a particular gene transcript is reflected in all of the RNAs or merely in the precursors.

5.5 SUMMARY AND DISCUSSION

Several systems exist for the study of mitochondrial biogenesis during higher plant development. The best characterised system is germination and early seedling development of V. faba (Dixon et al 1980, Forde and Leaver 1980) and this offers a model in which the causes of (i) a lag in the overall ability of mitochondria to synthesise protein and (ii) the specific synthesis of mitochondrial polypeptides during development can be investigated. However, the

experiments described in this chapter show that various factors have combined to make detection of mt transcripts in Northern blots of V. faba total cellular RNA difficult. Several approaches which might overcome these problems have been suggested.

In conclusion it seems likely that mitochondrial transcript analysis during germination and early development of V. faba will give an important insight into the factors controlling mitochondrial biogenesis in higher plants. The recent identification of higher plant nuclear genes encoding mitochondrial polypeptides (for example ATP-ADP translocator protein, A. Baker pers. commun. and F_1 ATPase β subunit, C. Leaver pers. commun.) should permit studies to be made on the co-ordinate expression of genes in the nucleus and the mitochondrion whose polypeptides form subunits of the same mosaic enzyme complexes.

CHAPTER 6 DISCUSSION

6.1 SUMMARY OF RESULTS AND AIMS OF THIS CHAPTER

The aim of the work described in this thesis was to increase our understanding of the expression of the higher plant mitochondrial genome with particular emphasis on the transcription of mitochondrial genes. While this objective has been achieved, the work presented in this thesis has posed far more questions than it has answered. In particular it has highlighted the need for much more detailed analysis before the way in which mitochondrial gene expression is controlled and co-ordinated with nuclear gene expression in the biogenesis of a functional mitochondrion is understood. The results of the two main lines of research which have been pursued in this thesis, namely a) the attempts to develop a cell free system for the transcription/translation of higher plant mt DNA and b) the characterisation of maize mitochondrial transcripts can be summarised as follows:

1. An E. coli cell free system fails to translate maize mitochondrial RNA. The addition of maize mt RNA to an E. coli S-30 does not significantly stimulate incorporation of ^{35}S methionine into protein but no novel polypeptides can be detected when analysed by SDS polyacrylamide gel electrophoresis and autoradiography. However the current availability of antibodies raised against purified maize mitochondrial polypeptides (for example F_1 ATPase α subunit, E. Hack pers. commun.) might permit identification of any mitochondrial polypeptides synthesised in low amounts in such a cell free system.

Possible reasons for the failure of the E. coli system to translate maize mt RNA were suggested. These included the absence of a ribosome

binding site 5' to higher plant mt genes showing homology to the bacterial Shine and Dalgarno sequence (Dawson et al 1984). Supplementation of an E. coli S100 (Modollel 1971) with higher plant mt ribosomes could overcome this problem.

2. Transcripts hybridising to four mitochondrial protein coding genes, namely COB, COI, COII and F_1 ATPase α subunit were identified showing that these genes (which had been characterised by sequence analysis only) are expressed in the mitochondrion. These represented the first plant mitochondrial protein coding gene transcripts to be unequivocally identified. Preliminary transcript analysis of the maize episomal DNAs S1 and S2 has revealed the presence in the mitochondrion of RNA species hybridising to DNA probes specific for S2 and the homologous region of S1 and S2.

3. Analysis of the transcripts of three maize mt protein coding genes (COB, COII and F_1 ATPase α subunit) has shown that all are transcribed in a far more complex manner than expected. Models to explain the origin of these multiple transcripts have been proposed.

4. The single copy gene (in N type mitochondria) encoding the maize COI polypeptide yields two major RNA species of 2.4 and 2.3 Kb. These RNAs differ in the length of their 5' untranslated regions which are ca 152 and 57 bases long. Analysis of the sequence around the start point of the higher molecular weight transcript (identified by S1 nuclease mapping) has revealed considerable sequence homology to a) the yeast nonanucleotide sequence thought to be involved in transcription initiation (Tabak et al 1983) and b) the sequences around the only other higher plant mitochondrial transcript precisely

mapped to date ie that of the Beta vulgaris mitochondrial 1.44 Kb plasmid (Munk-Hansen and Marcker 1984). Transcript analysis of the wheat COII transcript has also revealed homology to the yeast nona-nucleotide sequence (Bonen et al 1984). To establish whether these sequences represent part of a higher plant mitochondrial promoter will require more detailed transcription analysis of other higher plant mitochondrial genes together with identification of primary transcripts.

5. Transcripts of the maize COI, COII and F_1 ATPase α subunit genes have been studied in cytoplasmic male sterile C S and T type mitochondria. COI and F_1 ATPase α subunit transcripts in the male sterile cytoplasms studied and are also identical to N. In contrast COII transcripts in C S and T type mitochondria differ from each other and from N. The effects of genome rearrangements on the transcription of the COII gene will require more detailed sequence analysis and transcript mapping.

6. In common with certain yeast mt gene transcripts, the intron of the maize COII gene is excised from the precursor mRNAs and is possibly sufficiently stable to be detected by Northern blotting. In collaboration with A. Arnberg, University of Groningen, E.M. analysis of maize mt RNA has revealed the presence of circular RNAs of similar size to the COII intron, but the identity of these RNAs has not been confirmed.

7. Sequences homologous to the maize COII gene have been shown to be repeated in the maize mt genome but the exact nature of these homologous sequences has not been determined.

These findings have been extensively discussed in Chapters 3 and 4.

The purpose of the remainder of this chapter is to place these results in a broader context and to discuss some of the more speculative aspects of higher plant mitochondrial gene organisation and expression, and to make suggestions for future experiments.

6.2 GENERAL ASPECTS OF MITOCHONDRIAL GENE EXPRESSION IN HIGHER PLANTS

6.2.1 Multiple transcripts and the control of gene expression

Hybridisation of gene specific DNA probes to Northern blots of RNA isolated from four-day-old maize coleoptiles has shown that for all mt genes examined the patterns of transcription are much more complex than expected. In Chapter 4 various testable models have been proposed to account for these complex patterns of transcripts from each gene. Whether multiple transcripts arise from multiple transcription initiation events and/or multi-stop processing pathways is as yet unknown. However it is clear that the opportunities exist for the control of gene expression by modulating the amount of a mature translatable mRNA.

The generation of multiple transcripts from single copy genes is also found in the organelles of other organisms. In yeast mitochondria two tandem 'promoters' are found 5' to the gene encoding F_o ATPase subunit 9 (Edwards *et al* 1983a). Both are functional *in vivo* and *in vitro* although the most distal promoter sequence is the most efficient in each case. Multiple transcripts also hybridise to two yeast mt gene probes namely those for *COII* and F_o ATPase subunit 6, even though both of these genes are uninterrupted (Corruzzi *et al* 1981, van Ommen *et al* 1979). Thus it seems that processing (as opposed to splicing) is an important control point in the expression

of yeast mt genes. In maize chloroplasts, two transcripts are generated from the RuBPCase large subunit gene. Two transcripts hybridise to LS gene specific DNA probes. These differ at their 5' termini, these being separated by 238 bases (Crosland *et al* 1984). Both transcripts are synthesised in an *in vitro* transcription system, but only the higher molecular weight transcript can be 'capped' by guanylyl transferase. Thus the origin of the lower molecular weight transcript (ie whether it results from processing of the higher molecular weight transcript or is in fact a primary transcript possessing some unusual feature which inhibits capping) is unclear. Significantly the relative abundance of these two mRNA species changes during the light-induced development of the plastid. Both transcripts are present at all stages of development, but the higher molecular weight transcript increases in abundance during the first twenty hours of light-induced development and then decreases to etioplast levels by forty-four hours. It is not clear whether one or both transcripts are translatable *in vivo*. However, the discovery that the relative proportions of multiple transcripts of a particular gene vary at different stages of plastid biogenesis may be paralleled in higher plant mitochondria. Throughout the study presented in this thesis, only RNA isolated from four-day-old maize coleoptiles has been examined. It is possible that the proportions of various multiple transcripts change during mt biogenesis and in the transition from the etiolated seedling to the green plant. Clearly more detailed mapping of individual higher plant mt transcripts (ie the identification of primary and processed transcripts) together with transcript analysis at different stages of development is required before predictions of this sort can be made. Seemingly however, the opportunities for transcriptional and post transcriptional control of gene expression are

numerous in maize mitochondria. An important step towards a full understanding of gene expression in higher plant mitochondria will be to identify the translated mRNAs by isolation of polysomes and characterisation of the RNAs being translated.

The presence of multiple copies of two genes (two copies of α subunit in N type mitochondria and up to six copies of COI in S type mitochondria; A. Brennicke and P. Isaac per communs.) has important implications for the control of mitochondrial gene expression. It is possible for example that different copies are expressed at different stages of the life cycle of the higher plant. It is interesting that no differences between COI transcripts were detected in N and S type mitochondria. This suggests either that the additional gene(s) are not expressed or that their promoters and transcription terminator sequences are in identical positions with respect to the protein coding sequence. Sequence analysis of the flanking regions of the additional COI genes and/or in vitro transcription of individual genes will be necessary to determine whether they are expressed in S type mitochondria.

6.2.2 Polycistronic transcripts

There is now compelling evidence that many genes are expressed as part of polycistronic transcripts in human and S. cerevisiae mitochondria (Ojala et al 1981, Montoya et al 1982, Edwards et al 1983). In mammalian mitochondria, the genes are so closely spaced that reinitiation of transcription is unlikely to occur between them. Also, nascent and 'cappable' transcripts only map to one specific region of the mt genome in humans, suggesting that all transcripts are initiated from the same region.

In yeast mitochondria the evidence for polycistronic mRNAs is based on the following:

1. absence of the nonanucleotide sequence thought to be involved in transcription initiation 5' of some mt genes
2. inability to 'cap' transcripts of those genes lacking the nonanucleotide motif and
3. overlapping gene transcripts have been determined in some cases eg 21S RNA with $tRNA^{thr}$ and ATPase subunit 9 with $tRNA^{ser}$ (Miller *et al* 1983, Thalenfeld *et al* 1983).

Are higher plant mitochondrial genes co-transcribed? This is almost certainly true for the two higher plant mitochondrial 'cistrons' described below. In *Oenothera berteriana* mitochondria, an open reading frame preceding the *COII* gene shows structural analogy to the F_o ATPase subunit 8 genes in mammalian and fungal mitochondria. These genes overlap by four nucleotides and are co-transcribed as a bicistronic RNA species (Hiesel and Brennicke 1984a). In wheat mitochondria (Gray and Spencer 1983, Spencer *et al* 1984) the 18S and $tRNA^{f}_{met}$ genes are encoded on the same DNA strand and separated by one nucleotide. It is thus unlikely that transcription could be reinitiated between these genes. No transcripts hybridising to a precursor $tRNA^{f}_{met}$ - 18S rRNA have been detected by Northern blotting, but this could be due to very rapid processing of a primary transcript. In view of the proximity of the 5S rRNA to the 18S rRNA in wheat (114 bases, F. Quetier pers. commun.) it is possible that both 18S and 5S rRNAs are co-transcribed.

The discovery of a bicistronic transcript in *Oenothera berteriana* mitochondria and the likely co-transcription of $tRNA^{f}_{met}$ with 18S rRNA in wheat is surprising in view of the large size of the mitochondrial

genome of higher plants. This suggests there may be some selective advantage in compact grouping of genes which are co-transcribed. One possible advantage is that co-ordinate amounts of particular polypeptides and RNAs are synthesised. However in yeast this assumption may not be valid where the groups of genes under the control of one promoter are mixed ie protein coding genes, rRNAs and tRNAs are all co-transcribed.

Are protein coding genes in maize mitochondria co-transcribed? As yet there is no evidence for polycistronic transcripts. However, many of the transcripts are large enough to encode more than one gene (for example the ca 9 Kb COB transcript and the ca 6 Kb COII transcript). Sequence analysis 1500 bp 5' to COB shows that there are no open reading frames or tRNA-like structures in this region. Sequence analysis 5' to COII (ca 3000 bp) has shown that an ORF of at least 618 bases is fused to exon 1 of COII (Chapter 4, section 4.3.6.4). It is possible that this ORF is co-transcribed with the COII protein coding sequence. In this case the polypeptides would show considerable overlap, and the possibility that it represents a type of maturase has been discussed (Chapter 4, section 4.3.6.4). However incomplete data on the map position of COII transcripts makes any predictions of this sort difficult. The only fully mapped maize mt transcripts are those of COI, and sequence analysis 5' and 3' of the gene has revealed that no other open reading frames or tRNA-like structures are contained within either transcript. This of course does not preclude the possibility that COI transcripts form part of a longer rapidly processed primary transcript not detected by Northern blotting. Identification of primary transcripts and/or transcription initiation points in an in vitro transcription system

should resolve these questions. If maize mt genes are co-transcribed it is likely that the co-transcribed genes span large regions of DNA. For example COB in maize is separated from its known nearest putative gene (URF-1) by approximately 8 Kb (A. Dawson pers. commun. and Fig. 1.2). Although it is possible that other yet to be identified genes reside between COB and URF-1, extremely long primary transcripts are found in other organisms eg ca 90 Kb have been detected in Drosophila nuclei (Hogness pers. commun.).

6.3 PLANT MITOCHONDRIAL GENES AND GENOME ORGANISATION

6.3.1 The large size of the higher plant mitochondrial genome

Why is the plant mitochondrial genome so large in size? There are three possible explanations for the large size of the higher plant mitochondrial genome compared to other organisms. For example, the extra DNA could represent 1) additional genes or sequences necessary to mitochondrial function 2) DNA with a sequence independent function or 3) DNA that is non functional.

Does the higher plant mitochondrion contain genes not found in the mt genomes of other organisms? If so, how does this explain the variability in size detected between closely related plant species since closely related plants are likely to require similar mt gene complements (Ward et al 1981)? Analysis of the polypeptides synthesised by isolated maize mitochondria reveals the presence of ca 18 radioactively labelled products when separated on a one-dimensional SDS polyacrylamide gel. A further 20 - 30 can be resolved by two dimensional separation (Hack and Leaver 1983). The identification of the F_1 ATPase α subunit as a translation product of maize mitochonria (Hack and Leaver 1983) indicates that a gene not found in the

mitochondrial genomes of other organisms is present in higher plant mt DNA. However, the number of polypeptides synthesised is relatively constant between many higher plant species (E. Hack pers. commun.), suggesting that the presence of additional genes does not account for the variability in mt genome sizes among the higher plants.

It is possible that equating the number of translation products with the number of protein coding genes is not valid; for example premature termination of translation or proteolysis could lead to an overestimate of the number of genes while rare translation products may not be detected and hence the coding capacity underestimated. Even allowing for the capacity to encode 50 polypeptides, the large size of the higher plant mt genome is not accounted for. Variation in size is also found between different fungal mt genomes. Fungal mt genomes vary in size from the 19 Kb genomes of S. pombe (Lang et al 1983) and T. glabrata (Clark-Walker et al 1980) to the 108 Kb genome of Brettanomyces cursterii (Clark-Walker et al 1981). Some of this difference in size can be accounted for by a) long A + T rich spacer regions between genes and b) the presence of introns within genes. For example the S. pombe mt genome shows a very compact organisation with only short intergenic regions and three introns (Lang et al 1983). The additional DNA present in the higher plant mt genome is not due to the presence of A + T rich spacer regions as found in S. cerevisiae since the A + T content in higher plant mitochondria is relatively low (ca 53% in most species, Wells and Ingle 1970). Of course the presence of intergenic regions which are not A + T rich as found in N. crassa mt genome (66 Kb Burke and RajBhandary 1982) cannot be excluded. Whether the presence of introns contributes to the extra DNA in the higher plant mt genome is not known. The majority

of higher plant mt genes sequenced to date do not contain introns. In fact introns have only been found interrupting the COII genes from monocotyledonous plants (maize, Fox and Leaver 1981; wheat, Bonen et al 1984; rice, Kao et al 1984). The corresponding genes in dicotyledonous plants, pea (R. Wu pers. commun.) and Oenothera berteriana do not contain introns. This does not exclude the possibility that introns exist outside the protein coding sequence and as such are more difficult to detect. An intron is found 5' of a yeast nuclear gene encoding a ribosomal protein (Mitra and Warner 1984). The presence of such exogenic introns in higher plant mt DNA could be demonstrated by primer extension of mRNAs (purified by gel electrophoresis) and comparison of the sequence obtained with that of the genomic sequence.

If introns do not account for the extra mt DNA in higher plants then it must have a sequence independent function or even no function at all. In the latter case, the extra DNA persists in the mitochondrion because plants do not have the means to eliminate it. Animals in contrast have evolved the means for completely suppressing such extra DNA (or alternatively there are stronger selection pressures for animal mitochondria to retain the minimum possible mt DNA).

Two observations which may be relevant to the large size of the plant mt genome have been discussed in this thesis, namely 1) multiple copies of genes (two copies of ATPase α subunit in N type mitochondria and two 'major' COI copies in S type mitochondria) and 2) repeated fragments of genes (193 bp of COII in wheat and possibly part of the same gene is repeated in maize; Chapter 4, this thesis). The extra DNA may be partly explained by the duplication of genes and parts of genes. If these were subsequently scrambled by recombinational

events their origin would be obscured. It is possible that the repeated copies detected above may represent early stages of this process. Such events are known to occur in the mt DNA of Trypanosomes. 90% of the mt DNA of this organism consists of non-coding mini-circles with no known function (kinetoplast DNA, reviewed by Englund, 1981). Variations in the amount of kinetoplast DNA account for the size differences between two closely related Trypanosome species T. equiperdum (21 Kb) and T. brucei (320 Kb). Sequence analysis of two types of mini-circles in T. brucei revealed the presence of the same short sequence elements which were scrambled with respect to each other (Chen and Donaldson 1980).

If such events partly account for the large and variable size of the plant mt genome, such scrambling and amplification must have occurred after the evolutionary division of watermelon and muskmelon.

It is interesting to note that the mt DNA of the alga Chlamydomonas shows a compact organisation existing as a 16 Kb linear molecule with genes encoded on both strands (M. Gray pers. commun.).

6.3.2 The variable size of the higher plant mt genome

The size of the higher plant mt genome varies from 90 Kb in soyabean (Synenki et al 1978) up to 2400 Kb in muskmelon (Ward et al 1981). Within a single plant family, the Cucurbitaceae, mt genomes range in size from 330 Kb to 2400 Kb (Ward et al 1981). It is unlikely that these closely related species have significantly different mt gene coding requirements and the variable amount of DNA presents an enigma. A parallel situation is the presence of optional introns which are present in the mt genomes of some yeast strains but not others (for example COB introns 1, 2 and 3; Lazowska et al 1980, Foury and Tzagoloff 1976). These optional introns confer no obvious advantage

on the host strain. It has even been possible to construct COB genes which contain no introns by crossing wild type and petite strains. These genes all direct the synthesis of cytochrome b (although the lack of COB intron 4 prevents expression of COI in the absence of mim 2-1 and NAM 2-1 mutations; Labousse and Slonimski 1983, Dujardin et al 1982). Thus variation in the size of the mt genome between closely related organisms may be a general phenomenon.

In Chapter 1 the presence of chloroplast DNA sequences in the mt genome of higher plants was discussed. It is possible that the transfer of such sequences from other organelles to the mitochondrion accounts for the large and variable plant mt genome. However, although mt DNA from plants shows widespread homology to chloroplast DNA sequences, there is no apparent correlation between the size of a mitochondrial genome and the amount of chloroplast DNA it contains (Stern et al 1984). Hence the variability in the higher plant mt genome is not accounted for by import of chloroplast DNA, and the presence of nuclear DNA sequences in plant mitochondria has not been demonstrated.

Thus the origin of the extra DNA in the higher plant mt genome and the reasons why it is so variable in size remains an unanswered question. If it is due to the presence of large introns outside genes or within as yet unidentified genes then it will be identified by more detailed DNA sequence analysis and partial sequencing of transcripts. However, if the extra DNA has a sequence independent function or no function at all then in the absence of mt deletion mutants in higher plants, its origin and function may never be resolved.

6.4 IDENTIFICATION OF PLANT MITOCHONDRIAL GENES AND THEIR MUTANTS

As outlined in Chapters 1 and 3 novel approaches are likely to be necessary for the identification of genes in higher plant mitochondria. This is because

a) the large size of the higher plant mt genome is not amenable to direct sequence analysis

b) there is a lack of mitochondrial mutants among higher plants

c) a heterologous cell free system fails to translate mt DNA from higher plants with fidelity

d) the approach of identifying plant mt genes by heterologous hybridisation is limited by the availability of probes and their homology to plant mt DNA.

Novel ways of identifying plant mt genes might be as follows:

1. Purification of mitochondrially synthesised polypeptides followed by microsequencing would allow DNA oligonucleotides to be synthesised showing homology to the genes encoding them. These oligonucleotides could then be used to isolate a restriction fragment containing the gene (or part of the gene). Such an approach might be particularly useful in identifying the genes that encode the variant polypeptides in C S and T male sterile cytoplasms.

2. Open reading frames could be identified in a suitable expression vector (for example p ORF-1, Weinstock et al 1984).
Such an expression vector could also be used to synthesise large amounts of any polypeptides produced by that gene fragment, which in turn would permit antibody production and hence identification of the corresponding polypeptide. This approach has been used in the identification of the COB introns 2 and 4 encoded maturase polypeptides

in yeast mitochondria (Jacq et al 1984). Alternatively sequence analysis of ORFs will predict the amino acid sequence. From this synthetic polypeptides can be made which may also permit antibody production. Such an approach has been used successfully in the a) identification of maturase encoded proteins (COB introns 2 and 4 in yeast, Guiso et al 1984) and b) identification of polypeptides encoded by human mitochondrial open reading frames (Marriottini et al 1983).

The DNA rearrangements detected around certain genes in male sterile cytoplasms in maize (Chapter 4) may be useful model systems in the study of both higher plant mitochondrial gene structure and expression and the causes of cytoplasmic male sterility. Particularly interesting in this respect are the differences among COII transcripts in the various male sterile cytoplasms in maize. Complete transcription analysis of this gene might show a relationship between genome rearrangements and gene expression. This could then be extended to the study of cms. Thus for example the 13,000 M_r polypeptide characteristic of T cytoplasm (Forde et al 1979) may result from a defect in intron splicing from a pre mRNA with the result that a shortened polypeptide is produced by co-translation of an exon and part of an intron. Restorer genes might in this case act as a novel maturase imported into the mitochondrion from the cytoplasm, so allowing correct splicing and translation (analagous to the NAM-2-1 yeast nuclear mutation involved in correct splicing of COI intron 4 (Groudinsky et al 1981). Alternatively restorer genes might specifically suppress translation of this polypeptide. In yeast, nuclear gene products are known to be required for the translation of particular polypeptides (Muller et al 1984). A mutation in a restorer gene may prevent such a

'translation promoter' from being expressed.

Clearly however much more analysis of 'normal' mitochondrial gene expression together with the effects of DNA rearrangements on this is necessary before predictions about the causes of cytoplasmic male sterility in maize mitochondria can be made.

6.5 CONCLUDING REMARKS

It is clear from the work presented in this thesis that a full understanding of the biogenesis of a functional higher plant mitochondrion will take many years of detailed work and the development of novel techniques. Although the results presented in this thesis are preliminary in many respects it has laid the foundations for several potentially fruitful lines of research and highlighted the need for much more detailed analysis of higher plant mt gene structure, organisation and expression.

The major unresolved questions about higher plant mt gene expression and organisation are as follows:

1. Why are the transcripts of many plant mt genes seemingly so complex? Does this reflect control of mt gene expression by the use of multiple promoters, some of which may be active at different stages of the life cycle? Does it reflect control of gene expression by multiple processing steps of large and possibly polycistronic primary transcripts or does it reflect duplication and/or fragmentation of protein coding genes in the mt genome? Answers to these questions will require identification of primary transcripts and possibly the development of a homologous in vitro transcription system.

2. How is higher plant mt gene expression controlled during mitochondrial biogenesis? How is it co-ordinated with nuclear gene expression,

particularly in the formation of multimeric enzyme complexes composed of nuclear and mitochondrially encoded subunits?

3. Why is the plant mitochondrial genome so large and variable in size? What is the function of the extra DNA?

4. Why are chloroplast sequences found in plant mitochondria and mitochondrial sequences found in the nuclear DNA? Do they represent relics of a progenitor organelle or represent random acquisitions? If they are randomly acquired, why are they maintained?

5. Why do multiple copies of some genes exist in the mitochondria of higher plants? Are all copies expressed?

6. Why are some genes encoded in the mitochondrion of plants whereas the corresponding genes are encoded in the nucleus of other organisms? Does this reflect a transfer of genes from the mitochondrion to the nucleus over a long period of time such as occurs within the life cycle of Podospora anserina (Wright and Cummings 1983)? If so, how does such gene transfer occur?

Many of these questions are testable using established techniques - others require more ambitious approaches. Sequence analysis of mitochondrial genes is in itself useful and interesting but should be regarded only as a means to an end. It is the study of the expression of those genes particularly at different stages of the life cycle of the plant which is likely to shed most light on the biogenesis of this organelle.

Alwine, J.C., Kemp, D.J., and Stark, G.R., (1977)
Proc. Natl. Acad. Sci. USA. 74, 5350-5354

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H, Smith, A.J.H., Staden, R. and Young, I.G. (1981)
Nature 290, 457-465

Anderson, S., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982)
J. Mol. Biol. 156, 683-717

Anderson, W.F., Bosch, L., Cohn, W.E., Lodish, H., Merrick, W.C., Weissbach, H., Wittmann, H.G. and Wool, I.G. (1977)
FEBS Lett. 76 1-10

Anziano, P.Q., Perlman, P.S., Lang, B.F. and Wolf, K. (1983)
Curr. Genet. 7, 273-284

Arnberg, A.C., Van Ommen,G.J.B., Grivell, L.A., Van Bruggen, E.F.J. and Borst, P. (1980)
Cell 19 313-319

Attardi, G. (1981)
Trends Biochem. Sci. 6 86-89 and 100-103

Attardi, G., Cantatore, P., Chomyn, A., Crews, S., Gelfand, R., Merkel, C., Montoya, J. and Ojala, D. (1982)
In 'Mitochondrial Genes', Slonimski, P., Borst, P. and Attardi, G. (eds.)
Cold Spring Harbor Laboratory pp 51-71

Attardi, G., Doersen, C., Gaines, G., King, M., Montoya, J. and Shuey, D. (1983)
In 'Nucleo-Mitochondrial Interactions', Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp107-126

Baldacci, G. and Zennaro, E. (1982)
Eur. J. Biochem 127 411-416

Barrell, B.G., Bankier, A.T. and Drouin, J. (1979)
Nature 282,189-194

Becker, W.M., Leaver, C.J., Weir, E.M. and Riezman, H. (1978)
Plant Physiol 62, 542-549

Beckett, J.B. (1971)
Crop. Sci. 11, 724-727

Berk, A.J. and Sharp, P.A. (1977)
Cell 12, 721-732

Bernardi, G. (1982)
Trends Biochem. Sci. 7, 404-408

Bernardi, G. and Bernardi, G. (1980)
FEBS Lett. 115, 159-162

Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton D.A. (1981)
Cell 26, 167-180

Birnboim, H.C. and oly, J. (1979)
Nucl. Acids Res. 7, 1513-1523

Boedtker, H., (1968)
Methods Enzymol Vol. XII., part B pp 429-458

Boerner, P., Mason, T.L. and Fox, T.D. (1981)
Nucl. Acids Res. 9, 6379-6390

Bogenhagen, D.F., Applegate, E.F. and Yoza, B.A. (1984)
Cell 36, 1105-1113

Bogenhagen, D. and Clayton, D.A. (1974)
J. Biol. Chem. 249, 7991-7995

Bolivar, F. and Backman, K (1979)
Methods Enzymol. 68, 245-267

Bolivar, F. Rodriguez, R.L., Betlach, M.C. and Boyer, H.W.(1977)
Gene 2, 75-93

Bonen, L., Boer, P.H., and Gray, M.W. (1984)
Submitted to E.M.B.O.J.

Bonen, L. and Gray, M.W. (1980)
Nucl. Acids Res. 8, 319-335

Bonen, L., Huh, T.Y. and Gray, M.W. (1980)
FEBS Lett. 111, 340-346

Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E., Tzagoloff, A. and Macino, G. (1980)
J. Biol. Chem. 255, 11927-11941

Bonner, T.I., Brenner, D.J., Neufeld, B.R. and Britten, R.J. (1973)
J. Mol. Biol. 81, 123-135

Bonner, W.M. and Laskey, R.A. (1974)
Eur. J. Biochem 46, 83-88

Borst, P. (1972)
Ann. Rev. Biochem 41, 333-376

Borst, P. (1980)
In 'Proceedings of Mocbach Colloquia'
Bucher, Th., Sebald, W. and Weiss, H. (eds.)
Springer Berlag, Berlin pp 27-41

Borst, P. and Grivell, L.A. (1971)
see Borst, P. and Grivell, L.A. (1981).

Borst, P. and Grivell, L.A. (1978)
Cell 15, 705-723

Borst, P. and Grivell, L.A. (1981)
Nature 290, 443-444

Borst, P. and Grivell, L.A. (1981a)
Nature 289, 439-440

Borst, P. and Hoeijmakers, J.H.J. (1979)
Plasmid 2, 20-40

Bottomley, W., Higgins, T.J.V., and Whitfeld, P. (1976)
FEBS Lett. 63 120-124

Bottomley, W., and Whitfeld, P.R. (1979)
Eur. J. Biochem. 93 31-39

Bottomley, W. (1982)
In 'Methods in chloroplast molecular biology'; M. Edelman, R.B. Hallick and N-H Chua (eds.)
Elsevier, Amsterdam pp 573-580

Boutry, M., Briquet, M. and Goffeau, A. (1983)
J. Biol. Chem. 258, 8524-8526

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon P. (1978)
Proc. Natl. Acad. Sci. USA 75 4853-4857

Brennicke, A (1980)
Plant Physiol. 65, 1207-1210

Brettell, R.I.S., Thomas, E. and Ingram, D.S. (1980)
Theor. Appl. Genet. 58, 55-58

Broach, J.R. (1981)
In 'Molecular Biology of the yeast Saccharomyces cerevisiae'; Strathearn, J.N., Jones, E.W. and Broach, J.R., (eds.)
Cold Spring Harbor Laboratory, New York pp 445-470

Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978)
Proc. Natl. Acad. Sci. USA 75 4801-4805

Brosius, J., Dull, T.J. and Noller, H.F. (1980)
Proc. Natl. Acad. Sci, USA 77 201-204

Brown, T.A., Davies, R.W., Ray, J.A., Waring, R.B. and Sc azzocchio, C. (1983)
EMBO J. 2, 427-435

Brown, W.M. (1980)
Proc. Natl. Acad. Sci. USA 77, 3605-3609

Bruce, A.C. and Uhlenbeck, O.C. (1978)
Nucl. Acids Res. 5, 3665-3677

Burger, G. and Werner, S. (1983)
In 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 331-342

Burke, J.M. and RajBhandary, U.L. (1982)
Cell 31, 713-716

Cantatore, P. and Attardi, G. (1980)
Nucl. Acids Res. 8, 2605-2624

Casey, J. and Davidson, N. (1977)
Nucl. Acids Res. 4, 1539-1552

Cech, T. (1983)
Cell 34, 509-518

Chaconas, G. and van de Sande, J.H. (1980)
Methods Enzymol. 65, 75-85

Chamberlin, M.J. (1982)
In 'The Enzymes' Vol. XV Nucleic Acids Part B 3rd Edition ;
Boyer, P.D. (ed.)
Academic Press pp 61-86

Chang, D.D. and Clayton, D.A. (1984)
Cell 36, 635-643

Chang A.C.Y., Lansman, R.A., Clayton, D.A. and Cohen, S.N. (1975)
Cell 6, 231-244

Chao, S., Sederoff, R.R. and Levings, C.S. (1983)
Plant Physiol. 71, 190-193

Chen, K.K. and Donelson, J.E. (1980)
Proc. Natl. Acad. Sci. USA 77, 2445-2449

Christianson, T., Edwards, J., Levens, D., Locker, J. and Rabinowitz, M. (1982)
J. Biol. Chem. 257, 6494-6500

Christianson, T., Edwards, J.C., Mueller, D.M. and Rabinowitz, M. (1983)
Proc. Natl. Acad. Sci. USA 80, 5564-5568

Chuang, D.M. and Weissbach, H. (1973)
Arch. Biochem. Biophys. 157, 28-35

Church, G.M., Slonimski, P.P. and Gilbert, W. (1979)
Cell 18, 1209-1215

Clark-Walker, G.D. and Linnane, A.W. (1966)
Biochem. Biophys. Res. Comm. 25, 8-13

Clark-Walker, G.D., McArthur, C.R. and Daley, D.J. (1981)
Curr. Genet. 4, 7-12

Clayton, D.A. (1982)
Cell 28, 693-705
Clewell, D.B. (1972) J. Bacteriol. 110, 667-676
Coen, D.M. (1982)
In 'Methods in Chloroplast Molecular Biology'; M. Edelman, R.B. Hallick and N-H Chua (eds.)
Elsevier, Amsterdam pp 581-603

Cohen, S.N., Chang, A.C.Y., and Hsu, L. (1973)
Proc. Natl. Acad. Sci. USA 69, 2110-2114

Coruzzi, G., Bonitz, S.G., Thalenfeld, B.E. and Tzagoloff, A. (1981)
J. Biol. Chem. 256, 12780-12787

Cosson, J. and Tzagoloff, A. (1979)
J. Biol. Chem. 254, 42-43

Crossland, L.D., Rodermel, S.R. and Bogorad, L. (1984)
Proc. Natl. Acad. Sci. USA 81, 4060-4064

Dagert, M. and Ehrlich, S.D. (1979)
Gene 6, 23-28

Dale, R.M.K., Mendu, N., Ginsburg H. and Kridl, J.C. (1984)
Plasmid 11, 141-150

Dasgupta, R., and Kaesberg, P. (1982)
Nucl. Acids Res. 10, 703-713

Davies, J.W. and Verduin, B.J.M. (1979)
J. Gen. Virol. 44, 545-549

Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A. and Scazzocchio, C. (1982)
Nature 300, 719-724

Davies, R.W., Waring, R.B. and Brown T.A. (1983)
In 'Nucleo-Mitochondrial Interactions'; Schweyen, R.J., Wolf, K. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 179-189

Dawid, I.B. (1972)
Develop. Biology 29, 139-151

Dawson, A.J., Jones, V.P., and Leaver, C.J. (1984)
E.M.B.O. J. 3, 2107-2114

de Bruijn, M.H.L., Schreier, P.H., Eperon, I.C., Barrell, B.G., Chen, E.Y., Armstrong, P.W., Wong, J.F.H. and Roe, B.A. (1980)
Nucl. Acids Res. 8, 5213-5222

de Jonge, J.C. and de Vries, H. (1983)
Curr. Genet. 7, 21-28

Denhardt, D.T. (1966)
Biochem. Biophys. Res. Commun. 23, 641-646

de Roude, A., Van Loon, A.P.G.M. and Grivell, L.A. (1980)
Nature 287, 361-363

de Vries, H., de Jonge, J.C., Arnberg, A., Peijnenbruge, A.A.C.M. and Agsteribbe, E. (1983)
In 'Nucleo Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 343-356

de Zamaroczy, M., Marotta, R., Faugeron-Fonty, G., Goursot, R., Mangin, M., Baldacci, G. and Bernardi, G. (1981)
Nature 292, 75-78

Dhwale, S., Hanson, D.K., Alexander N.J., Perlman, P.S. and Mahler H.R. (1981)
Proc. Natl. Acad. Sci. 78, 1778-1782

Dieckmann, C.L., Koerner, T.J. and Tzagoloff, A. (1984)
J. Biol. Chem. 259, 4722-4731

Dieckmann, C.L., Tomlison, G., and Tzagoloff, A. (1984a)
J. Biol. Chem. 259, 4732 - 4738

Dixon, L.K., Forde, B.G., Forde, J. and Leaver, C.J. (1980)
In 'The Organisation and Expression of the Mitochondrial Genome'
Kroon, A.M. and Saccone, C. (eds.)
Elsevier, Amsterdam pp 365-368

Dixon, L.K. and Leaver, C.J. (1982)
Plant. Mol. Biol. 1, 89-102

Dubin, S.B., Benedek, G.B., Bancroft, F.C. and Freifelder D. (1970)
J. Mol. Biol. 54, 547-556

Dujardin, G., Jacq, C. and Slonimski, P.P. (1982)
Nature 298, 628-632

Dujardin, G., Labouesse, M., Netter, P. and Slonimski, P. (1983)
In 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 233-250

Duvick, D.N. (1965)
Adv. Genet. 13, 1-56

and Leaver CJ
Dyer, T.A (1981)
In The Biochemistry of plants Vol. 6 Marcus A. (ed) pp 112 - 168
Academic Press

Edwards, K. and Kossell, H. (1981)
Nucl. Acids Res. 9, 2853-2869

Edwards, J.C., Levens, D. and Rabinowitz, M. (1982)
Cell 31, 337-346

Edwards, J.C., Christianson, T., Mueller, D., Biswas, T.K., Levens, D., Li, D., Wettstein, J. and Rabinowitz, M. (1983)
In 'Nucleo Mitochondrial Interactions'; Schweyen, R.J., Wolf, K. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 69-78

Edwardson, J.R. (1970)
Bot. Rev. 36, 341-420

Ellis, R.J. (1976)
In 'Perspectives in Experimental Biology. Vol 2 Botany'; Sunderland, N. (ed.)
Pergamon Press pp 283-298

Ellis, R.J. (1982)
Nature 299, 678-679

Englund, P. (1981)
In 'Biochemistry and Physiology of Protozoa 2nd ed. Vol.4'; Levandowsky, M., Hunter,S.H. (eds.)
Academic Press, New York pp 333-383

Falcone (1983) see Schatz,G(1983)

Falconet, D., Lejeune, B., Quetier, F. and Gray, M.W. (1984)
E.M.B.O. J. 3, 297-302

Farrelly, F. and Butow, R.A. (1983)
Nature 301, 296-301

Fauron, C. M-R. and Wolstenholme, D.R. (1980)
Nucl. Acids Res. 8, 2439-2452

Faye, G. and Simon, M. (1983)
Cell 32, 77-87

Forde, B.G. and Leaver, C.J. (1980)
Proc. Natl. Acad. Sci. USA 77, 418-422

Forde, B.G., Oliver, R.J.C. and Leaver, C.J. (1978)
Proc. Natl. Acad. Sci. USA 75, 3841-3845

Forde, B.G., Oliver, R.J.C. and Leaver, C.J. (1979)
Plant Physiol. 63, 67-73

Foury, F., and Tzagoloff, A. (1976)
Mol. Gen. Genet. 149, 43-50

Fox, T.D. (1979)
Proc. Natl. Acad. Sci. USA 76, 6534-6538

Fox, T.D. and Leaver, C.J. (1981)
Cell 26, 315-323

Gabrielli, F. and Baglioni, C. (1977)
Nature 269, 529-531

Garriga, G., Bertrand, H. and Lambowitz, A. (1984)
Cell 36, 623-34

Gellissen, G., Bradfield, J.Y., White, B.N. and Wyatt, G.R. (1983)
Nature 301, 631-634

Gengenbach, B.G., Connelly, J.A., Pring, D.R. and Conde, M.F. (1981)
Theor. Appl. Genet. 59, 161-167

Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P. and Weissman, S.M. (1978)
J. Mol. Biol. 126, 813-846

Goddard, J.M. and Cummings, D.J. (1975)
J. Mol. Biol. 97 593-609

Godson, G.N. and Vapnek, D. (1973)
Biochem. Biophys. Acta. 299, 516-520

Geolet, P., Lomonossoff, G.P., Butler, P.J.G., Akam, M.E., Gait, M.J. and Karn, J. (1982)
Proc. Natl. Acad. Sci. USA 79, 5818-5822

Goldberg, M.L., Lefton, R.P., Stark, G.R. and Williams, J.G. (1979)
Methods. Enzymol. 68, 206-242

Grabowski, P.J., Zaug, A.J. and Cech, T.R. (1981)
Cell 23, 467-476

Gray, M.W. and Spencer, D.F. (1983)
FEBS Lett 161, 323-327

Greer, C.L. and Abelson, J. (1984)
Trends Biochem. Sci. 9, 139-141

Grierson, D. (1982)
In 'Gel Electrophoresis of Nucleic Acids: a practical approach';
Rickwood, D. and Hames, B.D. (eds.)
I.R.L. Press Ltd. Oxford pp 1-38

Grivell, L.A., Hensgens, L.A.M., Osinga, K.A., Tabak, H.F., Boer, P.H., Crusius, J.B.A., van der Laan, J.C., de Haan, M., van der Horst, G., Evers, R.F. and Arnberg, A.C. (1982)
In 'Mitochondrial Genes'; P. Slonimski, P. Borst and G. Attardi (eds.)
Laboratory pp 225-246
Cold Spring Harbor

Grivell, L.A., Tabak, H.F. and Borst, P. (1983)
In 'CRC Reviews in Biochemistry'
Vol.14 297-316

Grivell, L.A. (1983)
In 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 25-45

Groot, G.S.P., Mason, T.L. and van Harten-Loosbroek, N. (1979)
Mol. Gen. Genet. 174, 339-342

Groot, G.S.P., van Harten-Loosbroek, N., Van Ommen, G.J.B. and Pijst, H.L.A. (1981)
Nucl. Acids Res. 9, 6369-6377

Grossman, L. (1968)
Methods Enzymol. 12, 467-479

Groudinsky, O., Dujardin, G. and Slonimski, P.P. (1981)
Mol. Gen. Genet. 184, 493-503

Groudinsky, O., Carigani, G., Schiavon, E., Frezza, D., Bergantin, E. and Slonimski, P.P. (1983)
In 'Nucleo-Mitochondrial Interactions'; Schweyen, R.J., Wolf, K., and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 227-232

Guillemaut, P. and Weil, J.H. (1975)
Biochim. Biophys. Acta 407, 240-248

Guiso, N., Dreyfus, M., Siffert, O., Danchin, A., Spyridakis, A., Gargouri, A., Claisse, M. and Slonimski, P.P. (1984)
EMBO J 3, 1769-1772

Hack, E. and Leaver, C.J. (1983)
EMBO J 2, 1783-1789

Hack, E. and Leaver, C.J. (1984)
Curr. Genet. 8, 537-542

Hadler, H.I., Dimitrijevic, B. and Mahalingam, R.
Proc. Natl. Acad. Sci. USA 80, 6495-6499

Halbreich, A., Di Franco, A., Groudinsky, O., Casson, J. and Slonimski, P.P. (1975)
Biochem. Biophys. Res. Comm. 64, 1286-1292

Halbreich, A., Pajor, P., Foucher, M., Grandchamp, C. and Slonimski, P. (1980)
Cell 19, 321-329

Hartley, M.R., Wheeler, A. and Ellis, R.J. (1975)
J. Mol. Biol. 91, 67-77

Hauska, G., Hurt, E., Gabellini, N. and Lockau, W. (1983)
Biochim. Biophys. Acta. 726, 97-133

Hensgens, L.A.M., Grivell, L.A., Borst, P. and Bos, J.L. (1979)
Proc. Natl. Acad. Sci. USA 76, 1663-1667

Hensgens, L.A.M., Arnberg, A.C., Roosendaal, E., van der Horst, G., van der Veen, R., van Ommen, G-J.B. and Grivell, L.A. (1983)
J. Mol. Biol. 164, 35-58

Herrmann, R., Neugebauer, K., Pirkl, E., Zentgraff, H., and Schaller, H. (1980)
Mol. Gen. Genet. 177, 231-242

Hiesel, R. and Brennicke, A. (1983)
EMBO J 2, 2173-2178

Hiesel, R. and Brennicke, A. (1984)
submitted

Hiesel, R. and Brennicke. A. (1984a)
submitted

Hirsch, M. and Penman, S. (1974)
Cell 3, 335-339

Hollenburg, C.P., Borst, P. and van Bruggen, E.F.J. (1970)
Biochim. Biophys. Acta. 209, 1-15

Hopper, A., Furukawa, A.H., Pham, H.D., and Martin, N.C. (1982)
Cell 28, 543-550

Howe, C.J., Bowman, C.M., Dyer, T.A. and Gray, J.C. (1982)
Mol. Gen. Genet. 186, 525-530

Howe, C.J. Auffret, A.D., Doherty, A., Bowman, C.M., Dyer, T.A., and Gray, J.C. (1982a)
Proc. Natl. Acad. Sci. USA 79, 6903-6907

Howley, P.M., Israel, M.A., Law, M-F. and Martin, M.A. (1979)
J. Biol. Chem. 254, 4876-4883

Hu, N-t, and Messing, J. (1982)
Gene 17, 271-277

Hudspeth, M.E.S., Ainley, M.W., Shumard, D.S., Butow, R.A. and Grossman, L.I. (1982)
Cell 30, 617-626

Humphries, R.F., Ley, T., Turner, P., Davis Moulton, A. and Nienhaus, A.W. (1982)
Cell 30, 173-183

Iams, K.T. and Sinclair, J.H. (1982)
Proc. Natl. Acad. Sci. USA 79, 5926-5929

Isaac, P.G., Jones, V.P., and Leaver, C.J. (1984)
Submitted to Nucl. Acids Res.

Jacobs, H.T., Posakony, J.W., Grula, J.W., Roberts, J.W., Xin, J-H., Britten, R.J. and Davidson, E.H. (1983)
J. Mol. Biol. 165, 609-632

Jacq, C., Banroques, J., Becam, A.M., Slonimski, P.P., Guiso, N. and Danchin, A.
EMBO J 3, 1567-1572

Kao, T-h., Moon, E., and Wu, R. (1984)
Nucl. Acids. Res. 12, 7305-7316

Karabin, G.D., Farley, M. and Hallick, R.B. (1984)
Nucl. Acids. Res. 12, 5801-5812

Kemble, R.J. and Bedbrook, J.R. (1980)
Nature 284, 565-566

Kemble, R.J., Gunn, R.E. and Flavell, R.B. (1983)
Theor. Appl. Genet. 65, 129-144

Kemble, R.J. Mans, R.J., Gabay-Laughnan, S. and Laughnan, J.R. (1983a)
Nature 304, 744-747

Kemble, R.J. and Thompson, R.D. (1982)
Nucl. Acids Res. 10, 8181-8192

Kirby, K.S. (1968)
Methods Enzymol. 12, 87-99

Kiss, A., Sain, B., Venetianer, P. (1977)
FEBS Lett. 79, 77-79

Koller, B., Delius, H. and Dyer, T.A. (1982)
Eur. J. Biochem 122, 17-23

Koller, B., Gingrich, J.C., Stiegler, G.L., Farley, M.A., Delius, H. and Hallick, R.B. (1984)
Cell 36, 545-553

Kolodner, R. and Tewari, K.K. (1972)
Proc. Natl. Acad. Sci. USA 69, 1830-1834

Kroon, A.M. (1963)
Biochim. Biophys. Acta 72, 391-402

Kruger, R., Grabowski, P., Zaug, A., Sands, J., Gottschling, D. and Cech, T. (1982)
Cell 31, 147-157

Kuntzel, H. and Noll, H. (1967)
Nature 215, 1340-1345

Kuntzel, H. and Schafer, K.P. (1971)
Nature New Biol. 231, 265-269

Labouesse, M. and Slonimski, P.P. (1983)
EMBO J 2. 269-276

Laemmli, U.K. and Favre, M. (1973)
J. Mol. Biol. 80, 575-599

Lamb, A.J., Clark-Walker, G.D. and Linnane, A.W. (1968)
Biochim, Biophys. Acta. 161, 415

Lang, B.F., Ahne, F., Distler, S., Trinkl, H., Kaudewitz, F. and Wolf, K. (1983)
In 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 313-330

Langridge, P. (1981)
FEBS Lett. 123, 85-89

Laskey, R.A. and Mills, A.D. (1975)
Eur. J. Biochem. 56, 335-341

Laughnan, J.R. and Gabay, S.J. (1973)
Theor. Appl. Genet. 43, 109-116

Laughnan, J.R. and Gabay, S.J. (1975)
In 'Maize Genetics and Breeding'; Walden, D.B. (ed.)
Wiley, New York pp 427-446

Laughnan, J.R. and Gabay, S.J. (1975a)
In 'Genetics and Biogenesis of Mitochondria and Chloroplasts';
Birky, W.W. Jr., Perlman, P.S. and Byers, R.J. (eds.)
Ohio State University Press, Columbus pp 330-349

Lazarus, D.M. and Kuntzel, H. (1981)
Curr. Genet. 4, 99-107

Lazowska, J., Jacq, C. and Slonimski, P.P. (1980)
Cell 22, 333-348

Leaver, C.J. and Forde, B.G. (1980)
In 'Genome Organisation and Expression in Higher Plants'; Leaver, C.J. (ed.)
Plenum, New York pp 407-425

Leaver, C.J., Forde, B.G., Dixon, L.K. and Fox, T.D. (1982)
In 'Mitochondrial Genes'; Slonimski, P., Borst, P. and Attardi, G. (eds.)
Cold Spring Harbor Laboratory pp 457-470

Leaver, C.J. and Gray, M.W. (1982)
Ann. Rev. Plant Physiol. 33, 373-402

Leaver, C.J., Hack, E., Dawson, A.J., Isaac, P.G. and Jones, V.P. (1983)
In 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 269-284

Leaver, C.J., Hack, E., and Forde, B.G. (1983a)
Methods Enzymol. 97, 476-486

Leaver, C.J. and Harmey, M.A. (1973)
Biochem. Soc. Symp. 38, 175-193

Leaver, C.J. and Harmey, M.A. (1976)
Biochem. J. 157, 275-277

Leaver, C.J., and Ingle, J. (1971)
Biochem. J. 123, 235-243

Leaver, C.J. and Lovett, J.S. (1974)
Cell Differ. 3, 165-192

Leaver, C.J. and Pope, P.K. (1976)
In 'Nucleic Acids and Protein Synthesis in Plants'; Bogorad, L. (ed.)
Plenum Press, New York pp 213-237

Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977)
Biochemistry - USA 26, 4743-4751

Levens, D., Ticho, B., Ackerman, E. and Rabinowtiz, M. (1981)
J. Biol. Chem. 256, 5226-5232

Levens, D., Christianson, T., Edwards, J., Lustig, A., Ticho, B., Locker, J., and Rabinowitz, M. (1982)
In 'Mitochondrial Genes'; Slonimski, P.P., Borst, P., and Attardi, G. (eds.)
Cold Spring Harbor Laboratory, New York pp 295-310

Levings, C.S. III (1983)
Cell 32, 659-661

Levings, C.S. III, Kim, B.D., Pring, D.R., Conde, M.F., Mans, R.J. Laughnan, J.R. and Gabay-Laughnan, S.J. (1980)
Science 209, 1021-1023

Levings, C.S. III and Sederoff, R.R. (1983)
Proc. Natl. Acad. Sci. USA 80, 4055-4059

Levings, C.S. III, Shah, D.M., Hu, W.W.L., Pring, D.R. and Timothy, D.H. (1979)
In 'Extrachromosomal DNA, ICN-UCLA Symposium on Molecular and Cellular Biology'; Cummings, D.J., Borst, P., Dawid, I.B., Weissman, S.M., and Fox, C.F. (eds.)
Academic Press, New York pp 63-73

Lewis, M.K., and Burgess, R.R. (1982)
In 'The Enzymes' Vol. XV Nucleic Acids, Part B; Boyer P.D. (ed.)
Academic Press pp 110-150

Li, M. and Tzagoloff, A. (1979)
Cell 18, 47-53

Li, M., Tzagoloff, A., Underbrink-Lyon, K. and Martin, N.C. (1982)
J. Biol. Chem. 257, 5921-5928

Lodish, H.F. (1971)
J. Mol. Biol. 56, 627-632

Loening, U.E. (1967)
Biochem J. 102, 251-257

Lonsdale, D.M., Hodge, T.P., Fauron, C.M.-R., Flavell, R.B. (1983)
In 'Plant Molecular Biology' UCLA Symposia on Molecular Biology,
New Series Vol. 12
Goldberg, R.B. (ed.) pp 445-456

Lonsdale, D.M., Hodge, T.P., Howe, C.J. and Stern, D.B. (1983a)
Cell 34, 1007-1014

Luck, D.J.L. (1963)
J. Cell. Biol. 16, 483-499

Lund, E. and Dahlberg, J.E. (1979)
Proc. Natl. Acad. Sci. USA 76, 5840-5484

Luria, S.E. and Delbruck, M. (1943)
Genetics 28, 491-511

Macino, G. and Tzagoloff, A. (1979)
J. Biol. Chem. 254, 4617-4623

Macino, G. and Tzagoloff, A. (1980)
Cell 20, 507-517

Macreadie, I.G., Novitski, C.E., Maxwell, R.J., John, U., Ooi, B-G.,
McMullen, G.L., Lukins, H.B., Linnane, A.W. and Nagley, P. (1983)
Nucl. Acids. Res. 11, 4435-4451

Mandel, M. and Higa, A. (1970)
J. Mol. Biol. 53, 159-162

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982)
'Molecular Cloning, A Laboratory Manual'
Cold Spring Harbor Laboratory

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975)
Proc. Natl. Acad. Sci. USA 72, 1184-1188

Marmur, J. and Doty, P. (1962)
J. Mol. Biol. 5, 109-118

Marriottini, P., Chomyn, A., Attardi, A., Trovato, D., Strong, D.D.,
Doolittle, R.F. (1983)
Cell 32, 1264-1277

Marshall, J. (1952)
Phys. Rev. 86, 685-693

Martin, N.C., Miller, D.L., Najarian, D.R., Underbrink-Lyon, K.
In 'Mitochondrial Genes'; Slonimski, P., Borst, P., and Attardi, G. (eds.)
Cold Spring Harbor Laboratory pp 117-

Matthews, D.E., Gregory, P. and Gracen, V.E. (1979)
Plant Physiol. 63, 1149-1153

Maxam, A.M. and Gilbert, W. (1977)
Proc. Natl. Acad. Sci. USA 74, 560-564

McArthur, C.R. and Clark-Walker, G.D. (1983)
Curr. Genet. 7, 29-35

McConaughy, B.L., Laird, C.D. and McCarthy, B.J. (1969)
Biochemistry - USA 8, 3289-3295

McConkey, E.J. (1967)
Methods Enzymol. 12, 620-634

McDonnell, M.W., Simon, M.N. and Studier, F.W. (1977)
J. Mol. Biol. 110, 119-146

McIntosh, L., Poulsen, C., and Bogorad, L. (1980)
Nature 288, 556-560

McMaster, G.K. and Carmichael, G.G. (1977)
Proc. Natl. Acad. Sci. USA 74, 5350-5354

McNay, J.W., Pring, D.R. and Lonsdale, D.M. (1983)
Plant Mol. Biol. 2, 177-187

Merten, S., Synenki, R.M., Locker, J., Christianson, T. and Rabinowitz, M. (1980)
Proc. Natl. Acad. Sci. USA 77, 1417-1421

Messing, J. (1979)
Recombinant DNA Technical Bulletin,
NIH Publication No. 79-99, 2, No. 2, 43-48

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977)
Proc. Natl. Acad. Sci. USA 74, 3642-3646

Messing, J. and Vieira, J. (1982)
Gene 19, 269-276

Michel, F., Jacquier, A. and Dujon, B. (1982)
Biochimie 64, 867-881

Michel, F. and Dujon, B. (1983)
EMBO J. 2, 33-38

Miller, D. (1972)
Experiments in Bacterial Genetics
Cold Spring Harbor Laboratory, New York

Miller, D.L., Underbrink-Lyon, K., Najarian, D.R., Krupp, J., and Martin, N.C. (1983)
In 'Nucleo-Mitochondrial Interactions'
Schweyen, R.J., Wolf, K. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 151-164

Mitra, G. and Warner, J.R. (1984)
J. Biol. Chem. 259, 9218-9224

Modollel, J. (1971)
In 'Protein Synthesis in Bacterial Systems'; Last, J.A., and Laskin, A.I. (eds.)
M. Dekker, New York pp 2-58

Montoya, J., Ojala, D. and Attardi, G. (1981)
Nature 290, 465-470

Montoya, J., Christianson, T., Levens, D., Rabinowitz, M. and Attardi, G. (1982)
Proc. Natl. Acad. Sci. USA 79, 7195-7199

Montoya, J. Gaines, G.L. and Attardi, G. (1983)
Cell 34, 151-159

Moorman, A.F.M. and Grivell, L.A. (1976)
In 'Genetic Function of Mitochondrial DNA;' Saccone, C. and Kroon, A.M.
Elsevier/N. Holland Biomedical, Amsterdam

Moorman, A.F.M., Lamie, F., and Grivell, L.A. (1976)
FEBS Lett. 71, 67-72

Moorman, A.F.M., van Ommen, G.J.B. and Grivell, L.A. (1978)
Mol. Gen. Genet. 160, 13-24

Moorman, A.F.M., Verkley, F.N., Asselbergs, F.A.M. and Grivell, L.A. (1977)
In 'Mitochondria 1977': Genetics and Biogenesis of Mitochondria; Bandlow, W., Schweyen, R.J., Wolf, K. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 385-399

Morimoto, R., Locker, J., Synenki, R.M., and Rabinowitz, M. (1979)
J. Biol. Chem. 254, 12461-12470

Mukerjee, J. and Goldfeder, A. (1973)
Biochemistry USA 12, 5096-5101

Muller, P.P., Reif, M.K., Zonghou, S., Sengstag, C., Mason, T.L. and Fox, T.D. (1984)
J. Mol. Biol. 175, 431-452

Munk-Hansen, B.M. and Marcker, K.A. (1984)
Nucl. Acids Res. 12, 4749-4756

Munn, E.A. (1974)
The Structure of Mitochondria
Academic Press, London

Murphy, W.I., Attardi, B., Tu, C. and Attardi, G. (1975)
J. Mol. Biol. 99, 809-814

Nirenberg, M.W. and Matthei, J.H. (1961)
Proc. Natl. Acad. Sci. USA 47, 1588-1602

Nelson, M.A. and Macino, G. (1983)= Citterich, M.H., Morelli, G., Nelson, M.A., and Macino, G. (1983)
In 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and
Kaudewitz, F. (eds.)
De Gruyter, Berlin

Newman, D. and Martin, N[illegible] (1982)
Plasmid 7, 66-71

Nobrega, F.G. and Tzagoloff, A. (1980)
J. Biol. Chem. 255, 9828-9837

O'Brien, T.W. and Matthews, D.E. (1976)
In 'Handbook of Genetics, Vol. 5'; King, R.C. (ed.)
Plenum, New York pp 535-580

Oliver, N.A., Greenberg, B.D. and Wallace, D.C. (1983)
J. Biol. Chem. 258, 5834-5839

Ojala, D., Montoya, J. and Attardi, G. (1981)
Nature 290, 470-474

Osinga, K.A., De Haan, M., Christianson, T. and Tabak, H.F. (1982)
Nucl. Acids Res. 10, 7993-8006

Osinga, K.A. and Tabak, H.F. (1982)
Nucl. Acids Res. 10, 3617-3626

Osinga, K.A., de Vries, E., van der Horst, G.T.J. and Tabak, H.F. (1984)
Nucl. Acids Res. 12, 1889-1900

Osinga, K.A., de Vries, E., van der Horst, G. and Tabak, H.F. (1984a)
EMBO J 3, 829-834

Padgett, R.A., Mount, S.M., Steitz, J.A. and Sharp, P.A. (1983)
Cell 35, 101-102

Padmanaban, G., Hendler, F., Patzer, J., Ryan, R. and Rabinowitz, M. (1975)
Proc. Natl. Acad. Sci. USA. 72. 4293-4297

Palmer, J.D. (1983)
Nature 301, 92-93

Palmer, J.D. and Shields, C.R. (1984)
Nature 307, 437-440

Palmer, J.D., Shields, C.R., Cohen, D.B. and Orton, T.J. (1983)
Nature 301, 725-728

Parish, J.H., and Kirby, K.S. (1966)
Biochim. Biophys. Acta 129, 554-562

Paterson, B.M., Roberts, B.E., and Kuff, E.L. (1977)
Proc. Natl. Acad. Sci. USA 74, 4370-4374

Peattie, D.A. (1979)
Proc. Natl. Acad. Sci. USA 76, 1760-1764

Peebles, C.L., Gegenheimer, P. and Abelson, J. (1983)
Cell 32, 525-536

Petermann, M.L. Pavlovec, A., and Hamilton, G. (1972)
Biochemistry 11, 3925-3933

Pfisterer, J. and Butow, D.E. (1981)
Proc. Natl. Acad. Sci. USA 78, 4917-4921

Powling, A. (1981)
Mol. Gen. Genet. 183, 82-84

Pring, D.R., Conde, M.F., Schertz, K.F. and Levings, C.S.III (1982)
Mol. Gen. Genet. 186, 180-184

Pring, D.R. and Levings, C.S. III (1978)
Genetics 89, 121-136

Pring, D.R., Levings, C.S. III, Hu, W.W.L. and Timothy, D.H. (1977)
Proc. Natl. Acad. Sci. USA 74, 2904-2908

Pring, D.R. and Thornbury, D.W. (1975)
Biochim. Biophys. Acta 383, 140-146

Queen, C. and Rosenberg, M. (1981)
Cell 25, 241-249

Quetier, F. and Vedel, F. (1977)
Nature 268, 365-368

Quetier, F. and Vedel, F. (1980)
In 'Genome Organisation and Expression in Plants'; Leaver, C.J. (ed.)
Plenum, New York pp 401-406

Rabussay, D., Herrlich. P., Schweiger, M. and Zillig. W. (1969)
FEBS Lett. 4, 55-56

Radloff, R., Bauer, W. and Vinograd, J. (1967)
Proc. Natl. Acad. Sci. USA 57, 1514-1521

Randerath, K. (1970)
Anal. Biochem. 34, 188-205

Ravetch, J.V., Model, P. and Robertson, H.D. (1977)
Nature 265, 698-702

Rekosh, D.M.K., Russell, W.C., Bellet, A.J.D. and Robinson, A.J.
Cell 11, 283-296

Ricciardi, R.P., Miller, J.S. and Roberts, B.E.
Proc. Natl. Acad. Sci. USA 76, 4927-4931

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Bery, P. (1977)
J. Mol. Biol. 113, 237-251

Robberson, D., Aloni, Y., Attardi, G. and Davidson, N. (1971)
J. Mol. Biol. 60, 473-484

Roberts, B.E., Gorecki, M., Mulligan, R.C., Danna, K.J., Rozenblatt, S. and Rich, A. (1975)
Proc. Natl. Acad. Sci. USA 72, 1922-1926

Roberts, B.E. and Paterson, B.M. (1973)
Proc. Natl. Acad. Sci. USA 70, 2330-2334

Rodel, G., Holl, J., Schmelzer, C., Schmidt, C., Schweyen, R.J., Weiss-Brummer, B. and Kaudewitz, F. (1983)
In 'Nucleo-Mitochondrial Interactions'; Schweyen, R.J., Wolf, K. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 191-201

Sanders, J.P.M., Heyting, C., Verbett, M-Ph., Meijlink, F.C.P.W. and Borst, P. (1977)
Mol. Gen. Genet. 157, 239-261

Sanger, F., Nicklen, S. and Coulson, A.R. (1977)
Proc. Natl. Acad. Sci. USA 74, 5463-5467

Sanger, F., Coulson, A.R., Barell, B.G., Smith, A.J.H. and Roe, B.A. (1980)
J. Mol. Biol. 143, 161-178

Sanger, H.L., Ramm, K., Domdey, H., Gross, H.J., Henco, K. and Riesner, D. (1979)
FEBS Lett. 99, 117-122

Scazzocchio, C., Brown, T.A., Waring, R.B. and Davies, R.W. (1983)
In 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 303-312

Schardl, C.L., Londsale, D.M., Pring, D.R., and Rose, K.R. (1984a)
Nature 310, 292-295

Schardl, C.L., Pring, D.R., Fauron, C.M.-R., and Lonsdale, D.M. (1984b)
Submitted to Cell

Schatz, G. (1983)
In 'Biogenesis of Mitochondria' - University of Basel

Schnare, M.N. and Gray, M.W. (1982)
Nucl. Acids Res. 10, 3921-3932

Schreier, P.H. and Cortese, R. (1979)
J. Mol. Biol. 129, 169-172

Schroeder, R., Breitenbach, M., Schweyen, R.J. (1983)
Nucl. Acids Res. 11, 1735-1746

Schuster, A.M., Sisco, P.H. and Levings, C.S. III (1983)
In 'Plant Molecular Biology UCLA Symposium on Molecular and Cellular Biology' New Series, Vol. 12; R.B. Goldberg (ed.) pp 437-444

Schwarz, Z. and Kossel, H. (1980)
Nature 279, 520-522

Scragg, A.H. and Thomas, D.Y. (1975)
Eur. J. Biochem 56, 183-192

Seed, B. (1982)
Nucl. Acids Res. 10, 1799-1810

Sevarino, K.A. and Poyton, R.O. (1980)
Proc. Natl. Acad. Sci. USA 77, 142-146

Shah, D.M. and Levings, C.S. III (1974)
Crop. Sci. 14, 852-853

Shapiro, R. and Hachman, J. (1966)
Biochemistry USA 5, 2799-2807

Sharp, P.A., Berk, A.J. and Berget, J.M. (1980)
Methods Enzymol 65, 750-768

Sharp, P.A., Sugden, B. and Sambrook, J. (1973)
Biochemistry - USA 12, 3055-3063

Shine, J. and Dalgarno, L. (1974)
Proc. Natl. Acad. Sci. USA 71, 1342-1346

Shuman, S. and Hurwitz, J. (1982)
In 'The Enzymes 15'; P.D. Boyer (ed.)
Academic Press, New York pp 246-266

Silverthorne, J. and Ellis, R.J. (1980)
Biochim. Biophys. Acta 607, 319-330

Simon, M. and Gaye, G.
Proc. Natl. Acad. Sci. (USA) 81, 8-12

Schuter, A.M., Sisco, P.H. and Levings, C.S. III (1983)
In 'Plant Mol. Biol. UCLA Symp. on Mol. and Cell Biol. New Series vol 12 pp 437-444

Sisco, P.H., Garcia-Arenal, F., Zaitlin, M., Earle, E.A. and Gracen, V.E. (1984)
Plant Science Lett. 34 127-134

Smith, H.J. and Bogorad, L. (1974)
Proc. Natl. Acad. Sci. USA 71, 4839-4842

Smith, S.M. and Ellis, R.J (1981)
J. Mol. Appl. Genet. 1, 127-137

Soberon, X., Covarrubias, L. and Bolivar, F. (1980)
Gene 9, 287-305

Spencer, D.F., Bonen, L. and Gray, M.W. (1981)
Biochemistry USA 20, 4022-4029

Spencer, D.F., Schnare, M.N. and Gray, M.W. (1984)
Proc. Natl. Acad. Sci. USA 81, 493-497

Spruill, W.M. Jrd., Levings, C.S. III and Sederoff, R.R. (1980)
Dev. Genet. 1, 363-378

Steitz, J.A. and Jakes, K. (1975)
Proc. Natl. Acad. Sci. USA 72, 4734-4738

Stern, D.B., Dyer, T.A. and Lonsdale, D.M. (1982)
Nucl. Acids Res. 10, 3333-3340

Stern, D.B. and Lonsdale, D.M. (1982)
Nature 299, 698-702

Stern, D.B. and Newton, K. (1984)
Plant. Mol. Biol. Reporter 2, 8-15

Stern, D.B., and Palmer, J.D. (1984)
Proc. Natl. Acad. Sci. USA 81, 1946-1950

Stern, D.B. and Palmer, J.D. (1984a)
Submitted to Nucl. Acids Res.

Stevens, B. (1981)
In 'The Molecular Biology of the Yeast Saccharomyces. Life Cycle and Inheritance'; Strathern, J.N., Jones, E.W. and Broach, J.R. (eds.)
Cold Spring Harbor Laboratory pp 471-504

Strausberg, R.L.and Butow, R.A. (1977)
Proc. Natl. Acad. Sci. 74, 2715-2719

Studier, F.W. (1972)
Science 176, 367-376

Studier, F.W. and Rosenberg, A.H. (1981)
J. Mol. Biol. 153, 503-526

Suyama, Y. and Miura, K. (1968)
Proc. Natl. Acad. Sci. USA 60, 235-242

Synenki, R.M., Levings, C.S. III and Shah, D.M. (1978)
Plant Physiol. 61, 460-464

Tabak, H.F., Osinga, K.A., de Vries, E., van der Blick, A.M., van der Horst, G.T.J., Groot Koerkamp, M.J.A., van der Horst, G., Zwarthoff, E.C. and MacDonald, M.E. (1983)
In 'Nucleo-Mitochondrial Interactions'; Schweyen, R.J., Wolf, K. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 79-94

Tanaka, T. and Weisblum, B. (1975)
J. Bacteriol. 121, 354-362

Thalenfeld, B.E., Hill, J. and Tzagoloff, A. (1983)
J. Biol. Chem. 258, 610-615

Thomas, P.S. (1980)
Proc. Natl. Acad. Sci. USA 77, 5201-5205

Thomas, P.S. (1983)
In 'Methods Enzymol' 100, 255-265

Thompson, R.D., Kemble, R.J. and Flavell, R.B. (1980)
Nucl. Acids Res. 8, 1999-2008

Timmis, J.N. and Scott, N.S. (1983)
Nature 305, 65-67

Twigg, A.J. and Sherratt, D. (1980)
Nature 283, 216-218

Tzagoloff, A. (1982)
'Mitochondria'
Plenum Press, New York.

Ullstrup, A.J. (1972)
Ann. Rev. Phytopath. 10, 37-50

van den Boogart, P., van Dijk, S. and Agsteribbe, E. (1982)
FEBS Lett. 147, 97-100

van Ommen, G.-J.B., Groot, G.S.P. and Grivell, L.A. (1979)
Cell 18, 511-523

Vogt, V.M. (1973)
Eur. J. Biochem 33, 192-200

Walberg, M.W. and Clayton, D.A. (1983)
J. Biol. Chem. 258, 1268-1275

Walden, R. (1978)
Ph.D thesis, University of Edinburgh

Walden, R. and Leaver, D.J. (1981)
Plant Physiol. 67, 1090-1096

Walton, J.D. (1983)
Plant Mol. Biol. Reporter 1, 101-103

Ward, B.L., Anderson, R.S. and Bendich, A.J. (1981)
Cell 25, 793-803

Waring, R.B. and Davies, R.W. (1984)
Gene 28, 277-291

Warmke, H.E. and Lee, S.L.J. (1977)
J. Hered. 68, 213-222

Weaver, R.F. and Weissmann, C. (1979)
Nucl. Acids Res. 7, 1175-1193

Weber, K. and Osborn, M. (1969)
J. Biol. Chem. 224, 4406-4412

Weinstock, G.M., ap Rhys, C., Berman, M.L., Hampar, B., Jackson, D., Silhary, T.J., Weisemann, J., and Zweig, M. (1984)
Proc. Natl. Acad. Sci. USA 80, 4432-44

Weissinger, A.K., Timothy, D.H., Levings, C.S. III, Hu, W.W.L. and Goodman, M.M. (1981)
Maize Genet. Co-op. Newsletter 55, 86

Weissinger, A.K., Timothy, D.K., Levings, C.S. III and Goodman, M.M. (1983)
Genetics 104, 365-379

Wells, R. and Ingle, J. (1970)
Plant Physiol. 46, 178-179

Wesolowski, M. and Fukuhard, H. (1981)
Mol. Cell. Biol. 1, 387-393

Whitfeld, P.R. and Bottomley, W. (1983)
Ann. Rev. Plant Physiol. 34, 279-310

Widger, W.R., Cramer, W.A., Herrmann, R. and Trebst, A. (1983)
Submitted to Proc. Natl. Acad. Sci. USA

Wright, R.M. and Cummings, D.J. (1983)
Nature 302, 86-88

Wright, R.M., Horrum, M.A. and Cummings, D.J. (1982)
Cell 29, 505-515

Wu, G.J. and Dawid, I.B. (1972)
Biochemistry USA 11, 3589-3595

Yamamoto, K., Alberts, B.M., Benzinger, R., Lawhorne, L. and Teriber, G. (1970)
Virology 40, 734-744

Yin, S., Heckman, J., Sarnoff, J. and RajBhandary, U.L. (1980)
In 'Organisation and Expression of the Mitochondrial Genome'; Kroon, A.M. and Saccone, C. (eds.)
Elsevier, Amsterdam. pp 307-310

Yoza, B.K. and Bogenhagen, D.F. (1984)
J. Biol. Chem. 259, 3909-3915

Zassenhaus, H.P. and Butow, R.A. (1983)
In 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.)
De Gruyter, Berlin pp 95-106

Zassenhaus, H.P., Martin, N.C. and Butow, R.A. (1984)
J. Biol. Chem. 9, 6019-6027

Zelcer, A., Weaber, K.F., Balazs, E. and Zaitlin, M. (1981)
Virology 113, 417-427

Zoller, M.J. and Smith, M. (1982)
Nucl. Acids Res. 10, 6487-6500

Zurawski, G., Bohnert, H.J., Bottomley, W. and Whitfeld, P.R. (1983)
Proc. Natl. Acad. Sci. USA 79, 7699-7703

Zurawski, G., Bottomley, W.and Whitfeld, P.R. (1982)
Proc. Natl. Acad. Sci. USA 79, 6260-6264

Zurawski, G., Perot, B., Bottomley, W., Whitfeld, P.R. (1981)
Nucl. Acids Res. 9, 3251-3270

APPENDIX

PUBLICATIONS

Parts of this work have previously appeared or will be published in the following papers:

1) Leaver, C.J., Hack, E., Dawson, A.J., Isaac, P.G. and Jones V.P. (1983) in Mitochondria 1983 Schweyen, R.J., Wolf, K. and Kaudewitz, F. (eds)
De Gruyter, Berlin pp 269-284

2) Dawson, A.J., Jones, V.P., and Leaver, C.J. (1984)
E.M.B.O.J. 3 2107 - 2113

3) Isaac, P.G., Jones, V.P., and Leaver, C.J. (1984)
Submitted to Nucl. Acids Res.

4) Dawson A.J., Jones, V.P. and Leaver, C.J. (1985)
Submitted to Methods in Enzymology

5) Dawson, A.J., Isaac, P.G., Jones, V.P. and Leaver, C.J. (1985)
Submitted to Gene

The EMBO Journal vol.3 no.9 pp.2107–2113, 1984

The apocytochrome *b* gene in maize mitochondria does not contain introns and is preceded by a potential ribosome binding site

Adam J.Dawson[1], Valerie P.Jones and Christopher J.Leaver

Department of Botany, King's Buildings, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JH, UK
[1]Present address: Amersham International p.l.c., White Lion Road, Amersham, Bucks HP7 9LL, UK

Communicated by C.J.Leaver

The apocytochrome *b* (COB) gene has been isolated from maize (*Zea mays* L.) mitochondrial DNA. Sequence analysis reveals that the coding region of the gene is 1164 bp long and, in contrast with the homologous gene from yeast, does not contain introns or TGA (Trp) codons. The predicted polypeptide encoded by the gene has a mol. wt. of 42 868 daltons, and shows ~48% amino acid sequence homology with the corresponding yeast and mammalian polypeptides. Hydropathic profiles of the polypeptide indicate the presence of nine, membrane spanning hydrophobic domains suggesting that it is organised in the inner mitochondrial membrane in a similar fashion to that proposed for apocytochrome *b* in other organisms. The COB gene is preceded by a sequence 5′-AGTTGTCA-3′ which may act as a ribosome binding site in the mRNA since: (i) it shows 67.5% complementarity with an octanucleotide at the 3′ end of the maize mitochondrial 18S rRNA, located in a position homologous to that of the *Escherichia coli* Shine and Dalgarno sequence, and (ii) a similar sequence precedes several other plant mitochondrial genes at a distance of 15–20 nucleotides from the ATG initiation codon. RNA transcript analysis shows that the gene is transcribed in a complex manner with the presumed mature mRNA (~2.25 kb) probably being derived by sequential processing from a larger primary transcript.

Key words: apocytochrome *b*/maize/mitochondria/plant mtDNA

Introduction

The mitochondrial genomes of higher plants are considerably larger and more complex than comparable genomes in animal (14–18 kb) and fungal (19–94 kb) cells (Grivell, 1983). Current estimates suggest that the range extends from 218 kb in *Brassica* sp. (Palmer and Shields, 1984) to ~2400 kb in some cucurbits, with a 7- to 8-fold variation within a single plant family (Ward *et al.*, 1981). Analysis of plant mitochondrial DNAs (mtDNAs) by restriction enzyme digestion and renaturation kinetics suggests that the mitochondrial genomes are primarily composed of unique sequences carried on discrete circular molecules which are derived from each other by recombination. This intermolecular recombination is apparently mediated by short repeated sequences which may represent up to 5–10% of the total mtDNA (Palmer and Shields, 1984; Schardl *et al.*, 1984).

The large and variable size of the plant mitochondrial genome is not reflected in a variability in the base composition. The mtDNA of all plant species examined to date have a remarkably constant G+C content of ~47%, which together with sequence analysis, suggests that long stretches of A+T-rich non-coding DNA, such as are found in *Saccharomyces cerevisiae*, are absent.

These observations raise important questions as to why, for example, the plant mitochondrial genome is so large and, what is the functional significance, if any, of the wide range of genome size found in different plants. While the recent finding that mtDNA from a range of plant species show extensive and widespread homology to chloroplast DNA (Stern and Palmer, 1984) provides a partial answer to these questions, we are currently investigating whether plant mtDNA contains additional structural genes not present in other organisms. Analysis of the polypeptides synthesised by isolated plant mitochondria reveals the presence of at least 18–20 radioactively labelled polypeptides with mol. wts. ranging from 8000 to 58 000 (Leaver *et al.*, 1983a). The majority of these polypeptides are associated with the inner mitochondrial membrane and to date subunits I and II of cytochrome *c* oxidase and subunit 9 of the F_o portion of ATPase have been identified (Leaver and Gray, 1982). The recent demonstration that the α-subunit of the F_1 ATPase is synthesised in isolated plant mitochondria, and by extrapolation encoded in plant mtDNA, suggests that the plant mitochondrial genome does indeed contain additional genetic information, since this subunit is encoded in the nucleus in other organisms (Hack and Leaver, 1983; Boutry *et al.*, 1983).

The only genes so far characterised on the mtDNA of higher plants are those encoding the 26S, 18S and 5S rRNAs (the latter being unique to plant mtDNA; Leaver and Gray, 1982), and a single protein-coding gene encoding subunit II of cytochrome *c* oxidase (Fox and Leaver, 1981). This latter gene was initially identified by the use of heterologous hybridisation between maize (*Zea mays* L.) mtDNA restriction fragments and specific mitochondrial gene probes from yeast. Sequence analysis confirmed the identity of the cytochrome *c* oxidase subunit II gene (COII, formerly termed *moxI* by Fox and Leaver, 1981) by virtue of the homology of the predicted amino acid sequence with the corresponding yeast and bovine sequences. It also revealed major differences in gene structure and codon usage compared with the homologous gene in other organisms. The maize COII gene contains a unique, centrally located intron not found in the corresponding fungal and animal genes. In addition, it appears that plant mtDNAs exhibit a further variation in the once 'universal' genetic code. The CGG codon which normally specifies arginine probably codes for tryptophan, whereas the triplet UGA which encodes tryptophan in the fungal and animal mitochondrial genetic code is probably a stop codon, as in the 'universal' genetic code (Fox and Leaver, 1981; Hiesel and Brennicke, 1983).

To extend our knowledge of the information content, gene structure and expression of the plant mitochondrial genome, we have isolated the apocytochrome *b* gene (designated COB)

from maize. This gene has been sequenced in a number of mammalian (e.g., Anderson *et al.*, 1981; Bibb *et al.*, 1981) and fungal (e.g., Nobrega and Tzagoloff, 1980; Waring *et al.*, 1981) mitochondrial genomes, but not in any higher plant. In mammals the COB gene is relatively simple and does not contain introns, whereas in some fungi the gene has a mosaic organisation containing up to five introns in *S. cerevisiae* (Nobrega and Tzagoloff, 1980; Lazowska *et al.*, 1980; Jacq *et al.*, 1982). At least one of these introns, bI4 in yeast, encodes a protein ('maturase') which is thought to be involved in the excision of other introns (De la Salle *et al.*, 1982).

Here we describe the isolation and analysis of the gene for apocytochrome *b* from maize mitochondria. Our isolation of the COB gene provides the first evidence that apocytochrome *b*, a key component of the mitochondrial electron transport chain, is encoded in the mtDNA of higher plants. A comparison of the coding and surrounding sequences of the COB gene with those of COII suggest that TGA does not occur in plant mitochondrial genes. Furthermore, analysis of sequences 5′ to the COB gene has also revealed the presence of an octanucleotide sequence which shows complementarity to a sequence near the 3′ end of the maize 18S rRNA and could function as a ribosome binding site in maize mitochondrial messenger RNAs. The derived amino acid sequence for apocytochrome *b* has also allowed predictions to be made about the structure and function of this protein in the inner mitochondrial membrane (this paper, and Saraste, 1984).

Results

Identification and cloning of the maize COB gene

An M13 mp7 clone of a 750-bp *Mbo*I fragment from within the COB gene of *Kluyveromyces lactis* (M13-3.8: a gift from M.de Haan and L.A.Grivell, University of Amsterdam) was labelled with ^{32}P by second strand synthesis. This clone, which contains 65% of the coding sequence for apocytochrome *b*, was hybridised under non-stringent conditions ($T_m - 47°C$) to restriction fragments of maize mtDNA which had been separated by gel electrophoresis and transferred to a nitrocellulose filter. Major hybridisation was obtained with 1.8-kb *Hind*III, 5.1-kb *Eco*RI, 13.1-kb *Bam*HI and 10.5-kb *Sal*I fragments (Figure 1). The probe also hybridised to a number of additional mtDNA restriction fragments at much lower intensity. Essentially similar patterns of hybridisation were obtained using COB-specific probes from *S. cerevisiae*, *Aspergillus nidulans* and *Bos taurus* (data not shown).

The 1.8-kb *Hind*III fragment (Figure 1A,B) was isolated by electroelution from a 1% (w/v) agarose gel and cloned in the *Hind*III site of pBR328. A recombinant clone containing the 1.8-kb fragment was isolated and designated pZmH1790. Plasmid DNA from this clone was used to prepare a restriction map in order to guide subsequent DNA sequence analysis (included in Figure 2).

DNA sequence analysis confirmed that pZmH1790 contained a mtDNA fragment with a single long open reading frame (ORF) which was bounded at its 3′ end by a *Hind*III site. The ORF encoded 154 amino acids which showed good homology with the amino-terminal portion of published apocytochrome *b* sequences (e.g., yeast – Nobrega and Tzagoloff, 1980 and animals – Anderson *et al.*, 1981). To isolate the remainder of the gene, the 1.8-kb mtDNA fragment was isolated from pZmH1790 and used to probe a library of cloned *Eco*RI fragments of mtDNA. A single clone, designated pbE4, (see Figure 2) was identified and shown to contain a single 5.1-kb mtDNA fragment which co-migrated with the fragment identified by hybridisation with the *K. lactis* COB probe (Figure 1D). Similarly a clone designated pbB2 (see Figure 2), containing the 13.1-kb *Bam*HI fragment identified in Figure 1F was isolated from a library of cloned *Bam*HI mtDNA fragments.

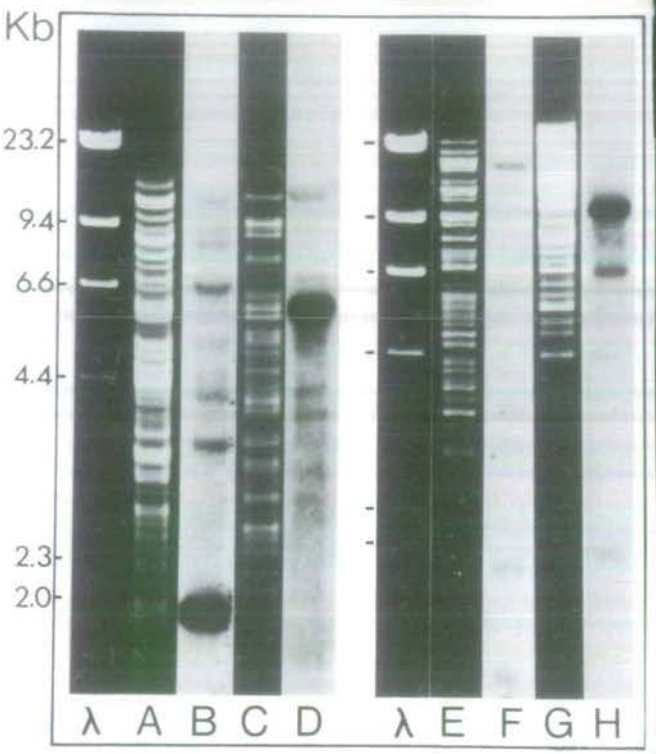


Fig. 1. Identification of maize mtDNA restriction fragments containing sequences homologous to the *K. lactis* gene encoding apocytochrome *b*. Maize mtDNA was digested with the restriction enzymes *Hind*III **(A,B)**, *Eco*RI **(C,D)**, *Bam*HI **(E,F)** and *Sal*I **(G,H)**, fractionated on 1% (w/v) agarose gels and stained with ethidium bromide **(A,C,E,G)**. DNA was transferred to nitrocellulose, hybridised with a ^{32}P-labelled *Mbo*I fragment of *K. lactis* mtDNA containing a portion of the COB gene and autoradiographed **(B,D,F,H)**. Mol. wt. marker, λ DNA digested with *Hind*III.

DNA sequence analysis

The three mtDNA fragments cloned in pZmH1790, pbE4 and pbB2 were digested with the restriction enzyme *Hind*III, *Eco*RI, *Sau*3A, *Msp*I, *Alu*I or *Taq*I and subcloned into M13 mp8 and mp9 (see Figure 2). DNA sequence analysis of these subclones by the dideoxy chain-termination method (Sanger *et al.*, 1980) revealed the presence of a continuous open reading frame of 1164 bp (Figure 3), which when translated, gave a predicted amino acid sequence with an overall homology of 48% and 47% with the corresponding yeast and beef sequences. The maize apocytochrome *b* protein deduced from the nucleotide sequence is 388 amino acids long and has a mol. wt. of 42 868 daltons.

Transcript analysis

To confirm that the maize COB gene was expressed *in vivo* and to investigate whether the initial gene transcript was processed, specific gene probes were hybridised to purified total mitochondrial RNA. The RNA was fractionated by electrophoresis on denaturing formaldehyde-agarose (1.3% w/v) gels, transferred to nitrocellulose and analysed by hybridisation with ^{32}P-labelled DNA probes. When the plasmid

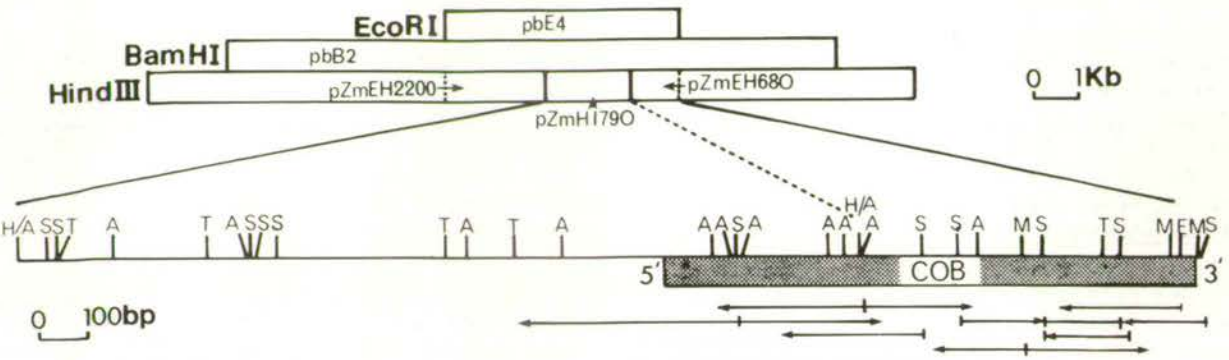


Fig. 2. Restriction map of maize mtDNA containing the apocytochrome *b* gene and flanking sequences. The pBR328 clones of mtDNA restriction fragments used in this study are indicated in the upper part of the figure. The lower portion of the figure indicates restriction sites used for M13 cloning and sequencing; the arrows below the map show the direction and extent of sequence analysis from each restriction site. Abbreviations: A, *Alu*I; B, *Bam*HI; E, *Eco*RI; H, *Hind*III; M, *Msp*I; S, *Sau*3A; T, TaqI.

```
-180                                *                                *
TTCGAGAATCAACTAACCAACAAATCCGTAGCCCAGGTGATTCGCTGCCTCCCTCTCGCCAAACCAAAATGGATGAATCTTCTCATGCAG

-90                                 *                                *
CTTTTTTCTTATTCAGGGCGCTGCGAAGCATCAAGGCAAGGGGGGTAAATAAAATAAGGGGGAAGAGGAGTTGTCACGATAGAAAAGAGAA

1                                   *                                *
ATGACTATAAGGAACCAACGATTCTCTCTTCTTAAACAACCTATATACTCCACACTTAACCAGCATTTAATAGATTATCCAACCCCGAGC
 M  T  I  R  N  Q  R  F  S  L  L  K  Q  P  I  Y  S  T  L  N  Q  H  L  I  D  Y  P  T  P  S
91                                  *                                *
AATCTTAGTTATTGGTGGGGGTTCGGTTGCTTAGCTGGTATTTGTTTAGTCATTCAGATAGTGACTGGCGTTTTTTTAGCTATGCATTAC
 N  L  S  Y  W  W  G  F  G  C  L  A  G  I  C  L  V  I  Q  I  V  T  G  V  F  L  A  M  H  Y
181                                 *                                *
ACACCTCATGTGGATCTAGCTTTCAACAGCGTAGAACACATTATGAGAGATGTTGAAGGGGGCTGGTTGCTCCGTTATATGCATGCTAAT
 T  P  H  V  D  L  A  F  N  S  V  E  H  I  M  R  D  V  E  G  G  W  L  L  R  Y  M  H  A  N
271                                 *                                *
GGGGCAAGTATGTTTCTCATTGTGGTTCACCTTCATATTTTTCGTGGTCTATATCATGCGAGTTATAGCAGTCCTAGGGAATTTGTTTGG
 G  A  S  M  F  L  I  V  V  H  L  H  I  F  R  G  L  Y  H  A  S  Y  S  S  P  R  E  F  V  W
361                                 *                                *
TGTCTCGGAGTTGTCATATTCCTATTAATGATTGTGACAGCTTTTATAGGATACGTACCACCTTGGGGTCAGATGAGCTTTTGGGGAGCA
 C  L  G  V  V  I  F  L  L  M  I  V  T  A  F  I  G  Y  V  P  P  W  G  Q  M  S  F  W  G  A
451                                 *                                *
ACAGTAATTACAAGCTTAGCTAGCGCCATACCAGTAGTAGGAGATACCATAGTGACTTGGCTTTGGGGTGGTTTCTCCGTGGACAATGCC
 T  V  I  T  S  L  A  S  A  I  P  V  V  G  D  T  I  V  T  W  L  W  G  G  F  S  V  D  N  A
541                                 *                                *
ACCTTAAATCGTTTTTTTAGTCTCCATCATTTACTCCCCCTTATTTTAGCAGGCGCCAGTCTTCTTCATCTGGCTGCATTGCATCAATAT
 T  L  N  R  F  F  S  L  H  H  L  L  P  L  I  L  A  G  A  S  L  L  H  L  A  A  L  H  Q  Y
631                                 *                                *
GGATCAAATAATCCATTGGGTGTACATTCTGAGATGGATAAAATTGCTTCTTACCCTTATTTTTATGTAAAGGATCTTGTAGGTCGGGTA
 G  S  N  N  P  L  G  V  H  S  E  M  D  K  I  A  S  Y  P  Y  F  Y  V  K  D  L  V  G  W  V
721                                 *                                *
GCTTCTGCTATCTTTTTTTCCATTTGGATTTTTTTTGCTCCAAATGTTTTGGGGCATCCCGACAATTATATACCTGCTAATCCGATGCCC
 A  S  A  I  F  F  S  I  W  I  F  F  A  P  N  V  L  G  H  P  D  N  Y  I  P  A  N  P  M  P
811                                 *                                *
ACCCCGCCTCATATTGTGCCGGAATGGTATTTCCTACCGATCCATGCCATTCTTCGCAGTATACCTGACAAAGCGGGGGGTGTAGCCGCA
 T  P  P  H  I  V  P  E  W  Y  F  L  P  I  H  A  I  L  R  S  I  P  D  K  A  G  G  V  A  A
901                                 *                                *
ATAGCACCAGTTTTTATATCTCTCTTGGCTTTACCTTTTTTTAAAGAAATGTATGTGCGTAGTTCAAGTTTTCGACCGATTCACCAAGGA
 I  A  P  V  F  I  S  L  L  A  L  P  F  F  K  E  M  Y  V  R  S  S  S  F  R  P  I  H  Q  G
991                                 *                                *
ATATTTTGGTTGCTTTTGGCGGATTGCTTACTACTAGGTTGGATCGGATGTCAACCTGTGGAGGCACCATTTGTTACTATTGGACAAATT
 I  F  W  L  L  L  A  D  C  L  L  L  G  W  I  G  C  Q  P  V  E  A  P  F  V  T  I  G  Q  I
1081                                *                                *
TCTTCTTTCTTTTTCTTCTTGTTCTTTGCCATAACGCCCATTCCGGGACGAGTTGGAAGAGGAATTCCAAAATATTACACGGAATAGACT
 S  S  F  F  F  F  L  F  F  A  I  T  P  I  P  G  R  V  G  R  G  I  P  K  Y  Y  T  E
1171
CATCGCACCGGATC
```

Fig. 3. DNA sequence of the maize mitochondrial COB gene and the predicted amino acid sequence of apocytochrome *b*. The predicted amino acid sequence was translated according to the universal genetic code except CGG was translated as tryptophan (W) as proposed by Fox and Leaver (1981). The proposed ribosome binding site preceding the ATG initiation codon is boxed.

pZmEH680 (see Figure 2) which contains sequences lying completely within the COB coding region, is used as a probe, a major transcript of ~2.25 kb is detected (Figure 4B). This is considerably larger than the 1164 bases which specify the coding region of the COB gene. In addition a less abundant transcript of ~4.2 kb and several minor transcripts of up to 9.0 kb are revealed on longer exposure of the blot. Hybridisation to the upstream clone pZmEH2200 (see Figure 2) reveals additional hybridisation to a number of high mol. wt. transcripts (Figure 4C) but does not identify the 2.25-kb transcript presumed to be the mature mRNA.

Identification of hydrophobic domains in the predicted apocytochrome b polypeptide

To provide further confirmation that the putative COB sequence encodes apocytochrome *b*, and also to predict the organisation of maize apocytochrome *b* in the inner mitochondrial membrane, a 'hydropathy' profile was constructed using the indices of Kyte and Doolittle (1982). Hydropathy values for blocks of 11 amino acids in the predicted sequence were calculated and the summed value plotted above the position of the central amino acid in the block (Figure 5). The block was successively displaced by one amino acid throughout the length of the polypeptide from amino to carboxy termini. Taking summed values for blocks in this way emphasises the presence of 'domains' with predominantly hydrophobic (positive hydropathy) or hydrophilic (negative hydropathy) characteristics, but reduces the effects of short term variation.

The plot identifies nine hydrophobic 'domains' in the polypeptide (I–IX), which are separated from each other by more hydrophilic regions. These domains correspond closely with the putative hydrophobic domains predicted by similar analyses of apocytochrome *b* in a variety of fungi and mammals, conducted by Saraste (1984) and by Widger *et al.*, 1984).

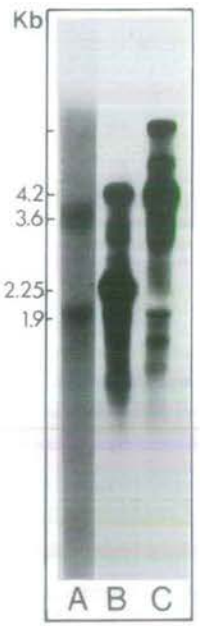


Fig. 4. Transcripts of the maize apocytochrome *b* gene. Total maize mtRNA was fractionated by electrophoresis through a 1.3% (w/v) agarose-formaldehyde gel and stained with ethidium bromide **(A)**. 26S and 18S mitochondrial RNAs are visualised as stained bands at ~3.6 kb and 1.9 kb, respectively. Parallel unstained tracks were transferred to nitrocellulose and hybridised with a ^{32}P-labelled **(B)** internal gene probe, pZmEH680 and with **(C)** an upstream probe, pZmEH2200 (see Figure 2 for origins of probes). *E. coli* rRNAs, cowpea chlorotic mottle virus and tobacco mosaic virus RNAs were used as RNA size standards.

Discussion

Cytochrome *b* is subunit IV of the multisubunit respiratory complex III (also called the cytochrome bc_1 complex of ubiquinol:cytochrome *c* oxidoreductase), located in the inner mitochondrial membrane of all organisms examined to date (Kreike *et al.*, 1979; Hauska *et al.*, 1983). Our determination of the maize COB gene sequence represents the first evidence that apocytochrome *b* is encoded in the plant mitochondrial genome.

Nucleotide sequence and codon usage

The single long ORF of 1164 bp (Figure 3) is assumed to represent the maize COB gene since it displays 52 and 47% sequence homology with the COB genes in *S. cerevisiae* (Nobrega and Tzagoloff, 1980) and *Homo sapiens* (Anderson *et al.*, 1981), respectively. The gene does not appear to contain introns, in contrast with COB in many fungi (Nobrega and Tzagoloff, 1980; Waring *et al.*, 1981; Citterich *et al.*, 1983) and the COII gene in maize mitochondria (Fox and Leaver, 1981). However, because of the low amino acid sequence conservation at the amino and carboxy termini of the predicted polypeptides (Figure 5, upper panel) the possibility that introns exist near the 5′ and 3′ ends of the gene cannot be formally excluded. Like the COB gene, the maize COI (Isaac *et al.*, in preparation) and *Oenothera* COII (Hiesel and Brennicke, 1983) genes do not contain introns.

The maize COB gene contains 12 TGG triplets, 11 at pos-

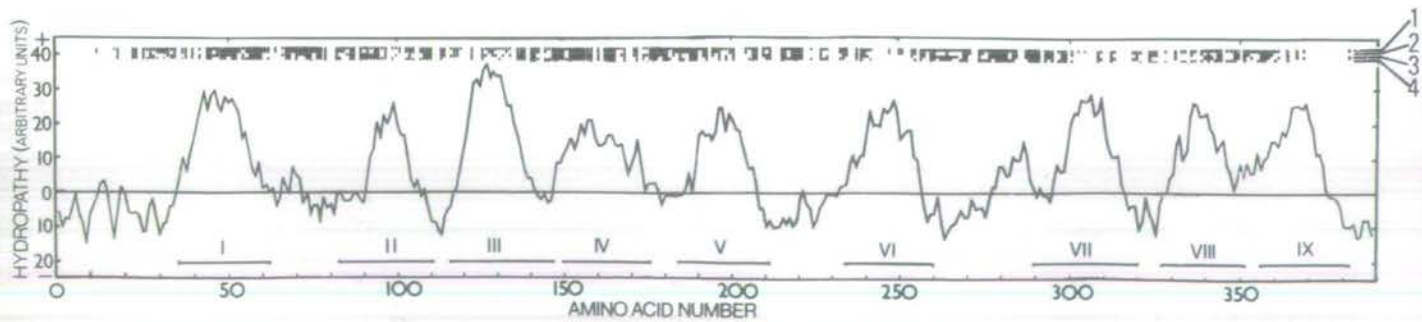


Fig. 5. Hydropathy profile of the predicted maize apocytochrome *b* polypeptide. The profile was calculated according to Kyte and Doolittle (1982) using an 11 amino acid window **(lower panel)**. Hydrophobic domains (I–IX) which are predicted to lie within the membrane have hydropathy >0, hydrophilic regions are <0. Amino acid 1 is encoded by nucleotides 1–3 shown in Figure 3. The upper bar diagram illustrates amino acid conservation between the predicted amino acid sequences of apocytochrome *b* from **(1)** maize, **(2)** *A. nidulans* (Waring *et al.*, 1981), **(3)** *S. cerevisiae* (Nobrega and Tzagoloff, 1980), and **(4)** *H. sapiens* (Anderson *et al.*, 1981). Identical amino acids are indicated by a solid vertical bar.

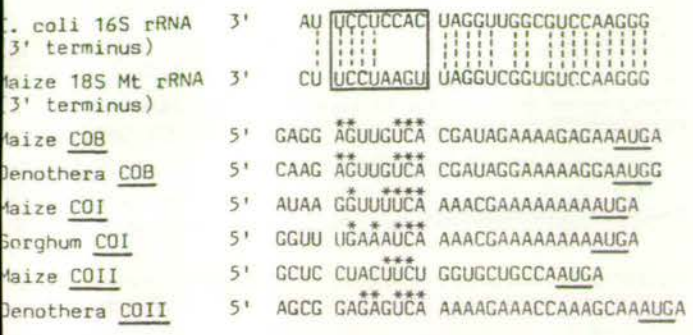


Fig. 6. A proposed ribosome binding site in higher plant mitochondrial mRNAs. The 3′-terminal sequences of *E. coli* 16S rRNA and maize 18S mt rRNA are compared, with identical nucleotides being shown by a vertical line. The Shine and Dalgarno sequence of the *E. coli* 16S rRNA, postulated to function in the attachment of the ribosome to mRNA, and the equivalent octanucleotide in maize 18S mt rRNA are boxed. Nucleotides 5′ to the initiation codon (underlined) of six plant mitochondrial genes which are capable of base pairing with the putative mRNA-binding sequence in maize mt rRNA are indicated with an asterisk. Source of sequence data: Maize: COB, this work; COI, P.G.Isaac, V.P.Jones and C.J.Leaver, in preparation; COII, Fox and Leaver, 1981. *Oenothera*: COB, A.Brennicke, personal communication; COII, Hiesel and Brennicke, 1983. Sorghum: COI, D.K.Hanson, J.N.Bailey-Serres, T.D.Fox and C.J.Leaver, unpublished data.

itions where the amino acid tryptophan is conserved at the homologous point in at least four of the eight other known apocytochrome *b* sequences. The triplet TGA is not found, however, which confirms that TGA is not used as a tryptophan codon in plant mitochondria (Fox and Leaver, 1981). A single CGG codon occurs in the gene (encoding amino acid 239), in a rather poorly conserved region of the polypeptide (Figure 5, upper panel). This triplet has been translated as Trp in view of the findings of Fox and Leaver (1981) and Hiesel and Brennicke (1983).

In the maize COB gene there is a consistent bias in favour of triplets ending in T (42%). A similar trend occurs in the maize COI and COII genes, and in the *Oenothera* COII gene, where on average 38% of the codons end in T. This contrasts with the bias in fungal mitochondrial genes, where T and A predominate (Waring *et al.*, 1981) and mammalian mitochondrial genes, where A and C are favoured in the third position (Anderson *et al.*, 1981).

Transcription of the maize COB gene

The major RNA species of 2.25 kb hybridising to the internal probe (pZmEH680; Figure 4B) is presumed to be the COB mRNA, because it is the most abundant transcript and because it does not hybridise to the upstream probe (pZmEH-2200). This mRNA is, however, ~1.1 kb longer than the COB gene (1164 bases) it encodes, suggesting that the molecule could carry coding sequences for other genes.

The transcript pattern identified by both the internal and upstream COB probes is strikingly complex. The use of strand-specific DNA probes shows that all of these hybridising RNA species arise from the COB encoding strand. The complexity of this pattern resembles that of the intron-containing maize COII gene (Fox and Leaver, 1981) but contrasts with that of the maize COI gene (V.P.Jones, unpublished data) which has no introns and apparently only two major transcripts. It is not known whether the various transcripts arise from multiple initiation and/or termination events, or from sequential processing of a longer precursor molecule. In yeast, the second alternative seems to apply (Osinga and Tabak, 1982; Christianson *et al.*, 1983), and the complexity of the observed transcript pattern depends as much on the relative rates of the processing steps as on the presence of introns (Coruzzi *et al.*, 1981; Thalenfeld *et al.*, 1983).

The origin of the two highest mol. wt. RNA species (Figure 4c) is of particular interest because the differential hybridisation of pZmEH680 and pZmEH2200 to these RNAs suggests that they could be transcripts of a gene upstream from COB, whose 3′ tail extends into the 5′ end of the COB gene. However, the 1600 bp of sequence 5′ to COB (shown partially in Figure 3) contains no open reading frames longer than 160 bp on either strand, nor any sequence capable of specifying conventional tRNAs.

A putative ribosome binding site upstream from plant mitochondrial initiation codons

The mRNAs encoded by higher plant mitochondria are seemingly much longer than the genes they contain, and it is likely that at least part of the additional sequence will lie 5′ to the initiation codon. Some mechanism must presumably exist, therefore, to confer specificity for translation initiation on the correct point in the mRNA. In the eubacteria (Rosenberg and Court, 1979) and chloroplasts (Whitfeld and Bottomley, 1983), RNA transcription starts 50–500 nucleotides upstream from the initiation codon of most protein coding genes. It has been postulated that specificity for translation initiation at this point is conferred by a ribosome binding site which is though to base pair with a complementary sequence (the Shine and Dalgarno sequence) located very near to the 3′ end of the 16S rRNA of the small ribosomal subunit (Shine and Dalgarno, 1974; Steitz and Jakes, 1975).

A comparison of the 18S rRNAs in wheat (Spencer *et al.*, 1984) and maize (Chao *et al.*, 1983) mitochondria with the *E. coli* 16S rRNA sequence, has revealed a remarkable conservation of primary and secondary structure (Spencer *et al.*, 1984). However, at the 3′ termini of the plant mitochondrial 18S rRNA sequences, part of the *E. coli* Shine and Dalgarno sequence (5′-CACCUCCU-3′) is specifically replaced by a novel sequence (5′-UGAAUCCU-3′) (Figure 6). Analysis of a variety of higher plant mitochondrial protein coding genes (Figure 6) shows that all appear to be preceded, ~13–18 nucleotides 5′ to the AUG initiation codon, by a tetranucleotide displaying 3/4 or 4/4 complementarity with the 18S rRNA-specific 'Shine and Dalgarno' type sequence, 5′-UGAA. In most, additional homology to the remainder of the 'Shine and Dalgarno' sequence occurs in this region. The only gene which displays poor complementarity is COII from maize, but recent sequencing results (Dawson *et al.*, unpublished data) indicate that translation of this gene may start several hundred nucleotides upstream from the sequence determined by Fox and Leaver (1981). The absence of the proposed ribosome binding site in the published sequence would not then be unexpected.

It remains to be established whether this sequence, potentially capable of base-pairing with the mitochondrial 18S rRNA, actually acts as a ribosome binding site. However, its ubiquity and constancy of position with respect to the initiation codon suggest that it could play some role in the initiation of translation. *E. coli* ribosome binding sites display a similar constancy of position, but in fungi the proposed mitochondrial ribosome binding site may fall anywhere between 8 and 116 nucleotides upstream from the initiation codon (Li *et al.*, 1982).

The sequences shown in Figure 6 also reveal that most

known higher plant mitochondrial AUG initiation codons are followed by A residues. Many prokaryotic genes are initiated with similar sequences, and it is thought that this A residue is involved in four base pair interactions with the $tRNA_f^{met}$ anticodon loop (Taniguchi and Weissman, 1978). Sequence analysis of the wheat mt $tRNA_f^{met}$ gene (Gray and Spencer, 1983) has shown that the anticodon loop contains the sequence $^{5'}$UCAU which would be capable of interacting with a $^{5'}$AUGA quadruplet in a similar fashion. It is possible, therefore, that this A residue could lend a second level of specificity to the AUG codon at which translation from the mRNA commences.

Higher order structures in the maize apocytochrome b polypeptide

The predicted polypeptide encoded by COB in maize contains 388 amino acids (a very similar number to apocytochrome *b* in other eukaryotes) and has a mol. wt. of 42 868. Preliminary models for the structure and functional organisation of this polypeptide in the inner mitochondrial membrane have recently been proposed (Saraste, 1984; Widger *et al.*, 1984), based on hydropathy plots (Kyte and Doolittle, 1982) of the predicted amino acid sequences of various fungal and animal apocytochromes *b*. Figure 5 (lower panel) shows a similar plot of the maize polypeptide, illustrating the presence of nine hydrophobic domains (labelled I to IX). This plot corresponds in remarkable detail to those of the bovine and yeast apocytochromes *b* (Saraste, 1984); hence, most of the numerous amino acid substitutions in the maize polypeptide (48% and 47% sequence homology to the yeast and human polypeptides, respectively) are likely to be silent in terms of their effects on protein conformation. Significantly, domain IX is preserved in both magnitude and position, despite a very low level of amino acid sequence homology between the various polypeptides in this region. This finding makes the existence of an intron in the 3′ region of the maize COB gene seem unlikely.

Saraste (1984) and Widger *et al.* (1984) suggest domains I–IX form hydrophobic membrane spanning α helices, and propose (based on the invariance of histidine residues 88, 102, 189 and 203) that the two electron-carrying protohaem prosthetic groups of cytochrome *b* are sandwiched in the membrane between helices II and V. The maize amino acid sequence could specify the requisite helices and contains the conserved histidine residues. Thus, if the models are correct, the protein encoded by the maize COB gene could form a biologically functional molecule.

Widger *et al.* (1984) have shown that some homology exists between the amino acid sequences of apocytochrome *b* from five mitochondrial sources, and two chloroplast encoded proteins: cytochrome b_6 (mol. wt. 23 000) and the '17-kd' protein of the b_6f complex. These two genes are adjacent on the chloroplast DNA map and the polypeptides can be aligned end-to-end with homologous sequences in the mitochondrial apocytochrome *b*. A gap of six amino acids separates the carboxy terminus of the aligned b_6 protein from the amino terminus of the aligned '17-kd' protein (Widger *et al.*, 1984). The maize apocytochrome *b* sequence also shows homology to the b_6 and '17-kd' protein sequences and, uniquely, contains a methionine residue which corresponds exactly with the position of the initiator methionine codon of the '17-kd' protein.

It is interesting to speculate that the sequence homologies between different functional genes in both chloroplasts and mitochondrial genomes, and the presence of chloroplast DNA sequences integrated into the plant mitochondrial genome, may be the result of a common ancestry or of an exchange of genetic information between these organelles during their evolution.

Materials and methods

Isolation of maize mitochondrial DNA and RNA

Seeds of maize (*Z. mays* L.) line B37N (Pioneer Hi-Bred International, Des Moines, IA) were grown in darkness at 28°C for 4 days. Gradient purified mitochondria were prepared from the coleoptile and mesocotyl tissue as described previously (Leaver *et al.*, 1983a) and mtDNA purified by solubilisation of the mitchondria in 1% (w/v) N-lauroyl sarcosine, 50 mM EDTA, 100 mM Tris-HCl (pH 8.0); digestion with 100 μg/ml proteinase K (Bethesda Research Labs.) at 60°C for 1 h followed by CsCl-EtBr equilibrium density gradient centrifugation (Fox, 1979).

Mitochondrial RNA was isolated from the purified mitochondria as described by Koller *et al.* (1982) and mtDNA removed by digestion with RNase-free DNase I (Worthington Biochemical Corp.) according to Smith and Ellis (1981).

Restriction enzyme digestion and gel electrophoresis of DNA

DNA (0.1 – 5 μg) was digested with restriction enzymes in 20 μl buffer under the conditions recommended by the suppliers (Bethesda Research Labs., Boehringer or Amersham International). mtDNA frequently contained a non-dialysable contaminant which inhibited digestions unless a 5- to 10-fold excess of restriction enzyme was used.

DNA fragments of 1–30 kb were separated by electrophoresis through 0.8 or 1% (w/v) agarose (Miles, type I low EEO) using a 40 mM Tris, 20 mM sodium acetate, 1 mM EDTA buffer, pH 8.2. Smaller fragments (10–1000 bp) were separated on 6% (w/v) polyacrylamide gels (29:1, acrylamide: bisacrylamide) using a 1 x TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3). DNA fragments were stained in ethidium bromide and recovered from agarose gels by electroelution (McDonell *et al.*, 1977) and from polyacrylamide gels by 'crush-soak' extraction (Maxam and Gilbert, 1977).

^{32}P-Labelling of DNA

Double-stranded DNA was labelled with [α^{32}P]dCTP by nick-translation (Rigby *et al.*, 1977). Single-stranded DNA clones in phage M13 were labelled by a modification of the sequencing reaction (Hu and Messing, 1982) except that the sequencing primer was used and strands were separated prior to hybridisation.

DNA:DNA hybridisation

DNA fragments were transferred to nitrocellulose as described by Southern (1975) and hybridised with ^{32}P-labelled probe under conditions determined by the expected degree of sequence homology between probe and blotted DNA (Howley *et al.*, 1979). These conditions, experimentally determined to yield the lowest signal:noise ratio, were: *homologous hybridisation*, pre-hybridisation (2 h), hybridisation (16 h) and first washes (3 x 20 min) in 4 x SSC, 1 x Denhardt's solution, 50% formamide and 25 μg/ml denatured herring sperm DNA (Serva) at 37°C. Final washes (4 x 5 min) in 2 x SSC at 20°C. Under these conditions, for maize mtDNA (G+C content 47%) hybridisation is at T_m – 26°C and the final wash at T_m – 38°C. *Heterologous hybridisation*, pre-hybridisation (2 h) and hybridisation (72 h) in 4 x SSC, 10 x Denhardt's solution, 0.1% (w/v) SDS and 100 μg/ml carrier DNA at 52°C. Washes (3 x 15 min) in 2 x SSC at 52°C. Hybridisation is at 47°C below the T_m of a maize mtDNA homoduplex and mtDNA sequences showing as little as 50% homology with the probe could be identified.

Recombinant DNA techniques

MtDNA restriction enzyme fragments were ligated to appropriately digested pBR328, M13 mp8 or M13 mp9 RFDNA (Messing and Vieira, 1982) using T4-DNA ligase (Boehringer). The ligated DNA was used to transform *E. coli* HB101 (plasmid vectors; Bolivar and Backman, 1979) or JM101 (M13 vectors, Messing and Vieira, 1982) as described by Dagert and Ehrlich (1979). *E. coli* colonies containing recombinant clones were selected either by antibiotic resistance characteristics (pBR328) or on the basis of digestion of the chromogenic substrate X-gal (M13). Plasmid clones containing sequences of interest were identified by colony hybridisation (Grunstein and Hogness, 1975); similarly M13 clones were identified by plaque hybridisation (Benton and Davis, 1977). pBR328 'libraries' of cloned mtDNA fragments generated after digestion with *Eco*RI, *Hind*III or *Bam*HI were prepared and maintained as described by Gergen *et al.* (1979).

DNA sequence analysis

All DNA sequence determination was performed by complementary strand elongation/termination of single-stranded template DNA cloned in M13 (Sanger *et al.*, 1977,1980) using [α^{32}P]dCTP (410 Ci/mmol, Amersham International). The sequencing strategies are shown in Figure 2.

RNA transcript analysis

RNA (10–20 μg) was fractionated under denaturing conditions in 1.3% (w/v) agarose-formaldehyde gels (Lehrach *et al.*, 1977) and blotted to nitrocellulose by the method of Thomas (1980). Parallel gel tracks of mtRNA and *E. coli* rRNA were stained with ethidium bromide to provide mol. wt. markers. Filters were pre-hybridised (4 h) and hybridised (16 h) in 50% (v/v) formamide, 5 x Denhardt's solution, 0.1% (w/v) SDS, 5 x SSC and 100 μg/ml herring sperm DNA at 42°C. Hybridisations contained 1–2 μg ^{32}P-labelled DNA (10^6–10^7 d.p.m./μg) and unbound probe was removed by two washes (10 min each) in 2 x SSC, 0.1% (w/v) SDS and two washes (10 min each) in 0.1 x SSC, 0.1% (w/v) SDS at 20°C.

Acknowledgements

We are grateful to Professor L.A.Grivell and Mr M.de Haan of the University of Amsterdam. Dr S.Anderson, M.R.C., Cambridge, and Dr T.D.Fox, Cornell University, for advice and the gift of heterologous gene probes. This work was supported by research grants from the Science and Engineering Research Council and the Agricultural Research Council to C.J.Leaver and postgraduate studentships to A.J.Dawson and V.P.Jones.

References

Anderson,S., Bankier,A.T., Barrell,B.G., De Bruijn,M.H.L., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., Schreier, P.H., Smith,A.J.H., Staden,R. and Young,I.G. (1981) *Nature,* **290**, 457-464.

Benton,W.D. and Davis,R.W. (1977) *Science (Wash.),* **196**, 180-182.

Bibb,M.J., Van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) *Cell,* **26**, 167-180.

Bolivar,F. and Backman,K. (1979) *Methods Enzymol.,* **68**, 245-267.

Boutry,M., Briquet,M. and Goffeau,A. (1983) *J. Biol. Chem.,* **258**, 8524-8526.

Chao,S., Sederoff,R.R. and Levings,C.S., III. (1983) *Plant Physiol.,* **71**, 190-193.

Christianson,T., Edwards,J.C., Mueller,D.M. and Rabinowitz,M. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 5564-5568.

Citterich,M.H., Morelli,G. and Macino,G. (1983) *EMBO J.,* **2**, 1235-1242.

Coruzzi,G., Bonitz,S.G., Thalenfeld,B.E. and Tzagoloff,A. (1981) *J. Biol. Chem.,* **256**, 12780-12787.

Dagert,M. and Ehrlich,S.D. (1979) *Gene,* **6**, 23-28.

De la Salle,H., Jacq,C. and Slonimski,P.P. (1982) *Cell,* **28**, 721-732.

Fox,T.D. (1979) *J. Mol. Biol.,* **130**, 63-82.

Fox,T.D. and Leaver,C.J. (1981) *Cell,* **26**, 315-323.

Gergen,J.P., Stern,R.H. and Wensik,P.C. (1979) *Nucleic Acids Res.,* **7**, 2115-2136.

Gray,M.W. and Spencer,D.F. (1983) *FEBS Lett.,* **161**, 323-327.

Grivell,L.A. (1983) in Wolf,K., Schewyen,R.J. and Kaudewtiz,F. (eds.), *Mitochondria 1983,* De Gruyter, Berlin NY, pp. 25-45.

Grunstein,M. and Hogness,D.S. (1975) *Proc. Natl. Acad. Sci. USA,* **72**, 3961-3965.

Hack,E. and Leaver,C.J. (1983) *EMBO J.,* **2**, 2783-2789.

Hauska,G., Hurt,E., Gabellini,N. and Lockau,W. (1983) *Biochim. Biophys. Acta,* **726**, 97-133.

Hiesel,R. and Brennicke,A. (1983) *EMBO J.,* **2**, 2173-2178.

Howley,P.M., Israel,M.A., Law,M.-F. and Martin,M.A. (1979) *J. Biol. Chem.,* **254**, 4876-4883.

Hu,N-t and Messing,J. (1982) *Gene,* **17**, 271-277.

Jacq,C., Pajot,P., Lazowska,J., Dujardin,G., Claisse,M., Groudinsky,O., de la Salle,H., Grandchamp,C., Labouesse,M., Gargouri,A., Guiard,B., Spyridakis,A., Dreyfus,M. and Slonimski,P.P. (1982) in Slonimski,P.P., Borst,P. and Attardi,G. (eds.), *Mitochondrial Genes,* Cold Spring Harbor Laboratory Press, NY, pp. 155-183.

Koller,B., Delius,H. and Dyer,T.A. (1982) *Eur. J. Biochem.,* **122**, 17-23.

Kreike,J., Bechmann,H., van Hemert,F.J., Schewyen,R.J., Boer,P.H., Kaudewitz,F. and Groot,G.S.P. (1979) *Eur. J. Biochem.,* **101**, 607-617.

Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.,* **157**, 105-132.

Lazowska,J., Jacq,C. and Slonimski,P.P. (1980) *Cell,* **22**, 333-348.

Leaver,C.J. and Gray,M.W. (1982) *Annu. Rev. Plant Physiol.,* **33**, 373-402.

Leaver,C.J., Hack,E. and Forde,B.G. (1983a) *Methods Enzymol.,* **43**, 476-484.

Leaver,C.J., Hack,E., Dawson,A.J., Isaac,P.G. and Jones,V.P. (1983b) in Wolf,K., Schewyen,R.J. and Kaudewitz,F. (eds.), *Mitochondria 1983,* De Gruyter, Berlin-NY, pp. 269-283.

Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry (Wash.),* **26**, 4743-4751.

Li,M., Tzagoloff,A., Underbrink-Lyon,K. and Martin,N.C. (1982) *J. Biol. Chem.,* **257**, 5921-5928.

Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 560-564.

McDonnell,M.W., Simon,M.N. and Studier,F.W. (1977) *J. Mol. Biol.,* **110**, 119-146.

Messing,J. and Vieira,J. (1982) *Gene,* **19**, 269-276.

Nobrega,F.G. and Tzagoloff,A. (1980) *J. Biol. Chem.,* **255**, 9828-9837.

Osinga,K.A. and Tabak,H.F. (1982) *Nucleic Acids Res.,* **10**, 3617-3626.

Palmer,J.D. and Shields,C.R. (1984) *Nature,* **307**, 437-440.

Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Bery,P. (1977) *J. Mol. Biol.,* **113**, 237-251.

Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.,* **13**, 319-353.

Sanger,F., Air,G.M., Barrell,B.G., Brown,N.L., Coulson,A.R., Fiddes,J.C., Hutchinson,C.A., Slocombe,P.M. and Smith,M. (1977) *Nature,* **265**, 687-695.

Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.,* **143**, 161-178.

Saraste,M. (1984) *FEBS Lett.,* **166**, 367-372.

Schardl,C.L., Lonsdale,D.M., Pring,D.R. and Rose,K.R. (1984) *Nature,* in press.

Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA,* **71**, 1342-1346.

Smith,S.M. and Ellis,R.J. (1981) *J. Mol. Appl. Genet.,* **1**, 127-137.

Southern,E.M. (1975) *J. Mol. Biol.,* **98**, 503-517.

Spencer,D.F., Schnare,M.N. and Gray,M.W. (1984) *Proc. Natl. Acad. Sci. USA,* **81**, 493-497.

Steitz,J.A. and Jakes,K. (1975) *Proc. Natl. Acad. Sci. USA,* **72**, 4734-4738.

Stern,D.B. and Palmer,J.D. (1984) *Proc. Natl. Acad. Sci. USA,* in press.

Taniguchi,T. and Weissman,C. (1978) *J. Mol. Biol.,* **118**, 533-565.

Thalenfeld,B,E., Hill,J. and Tzagoloff,A. (1983) *J. Biol. Chem.,* **258**, 610-615.

Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 5201-5205.

Ward,B.L., Anderson,R.S. and Bendich,A.J. (1981) *Cell,* **25**, 793-803.

Waring,R.B., Davies,R.W., Lee,S., Grisi,E., McPhail-Berks,M. and Scazzocchio,C. (1981) *Cell,* **27**, 4-11.

Whitfeld,P.R. and Bottomley,W. (1983) *Annu. Rev. Plant Physiol.,* **34**, 279-310.

Widger,W.R., Cramer,W.A., Herrmann,R. and Trebst,A. (1984) *Proc. Natl. Acad. Sci. USA,* **81**, 674-678.

Received on 25 May 1984; revised on 20 June 1984